

QY 1801 GAGTGGATACCATGTTTCAAAATCCGAAATACTGGAATCCTGGAGATGAAGTTGGTTC 1860
Db 1899 GAGTGGATACCATGTTTCAAAATCCGAAATACTGGAATCCTGGAGATGAAGTTGGTTC 1958
QY 1861 AAAATGAATCAAAACCAACGATTCGTTGGAGGAATGGAGCATTTGATATGCTGCA 1920
Db 1959 AAAATGAATCAAAACCAACGATTCGTTGGAGGAATGGAGCATTTGATATGCTGCA 2018
QY 1921 CTGGGATTTAG 1932
Db 2019 CTGGGATTTAG 2030

RESULT 2
U80032/c
LOCUS U80032 45389 bp DNA linear INV 16-SEP-2004
DEFINITION Caenorhabditis elegans cosmid C32E12, complete sequence.
ACCESSION U80032
VERSION U80032.1 GI:2073526
KEYWORDS HTG.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 45389)

REFERENCE
AUTHORS WormBase Consortium
CONSTRM Genome sequence of the nematode C. elegans: a platform for
TITLE investigating biology. The C. elegans Sequencing Consortium
JOURNAL Science 282 (5396), 2012-2018 (1998)
MEDLINE 99069613
PUBMED 9851916
REFERENCE 2 (bases 1 to 45389)
AUTHORS Wilcox,L.
TITLE The sequence of C. elegans cosmid C32E12
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 45389)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (27-NOV-1996) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
4 (bases 1 to 45389)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (08-MAY-1997) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
5 (bases 1 to 45389)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (18-OCT-2001) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
6 (bases 1 to 45389)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (19-APR-2002) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
7 (bases 1 to 45389)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2002) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
8 (bases 1 to 45389)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (19-NOV-2002) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
9 (bases 1 to 45389)

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
CONSTRM
TITLE
JOURNAL
COMMENT

Waterston,R.
Direct Submission
Submitted (13-JAN-2003) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
10 (bases 1 to 45389)
WormBase Consortium
Direct Submission
Submitted (16-SEP-2004) Department of Genetics, Washington
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
Louis, MO 63110, USA
On May 8, 1997 this sequence version replaced gi:1703546.
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University
St. Louis , MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
email: submissions@watson.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate chemistry
or covered by high quality data (i.e., phred quality >= 30); an
attempt was made to resolve all sequencing problems, such as
compressions and repeats; all regions were covered by sequence from
more than one m13 subclone.

For a graphical representation of this clone sequence and its
analysis see:
<http://www.wormbase.org/db/seq/sequence?name=C32E12;class=Sequence>

NEIGHBORING CLONE INFORMATION

The 5' clone is F56A3, 200 bp overlap; the 3' clone is B0261, 3201
bp overlap. Actual start of this clone is at base position 1 of
C32E12; actual end is at 42388 of C32E12.

NOTES:

Coding sequences below are the result of integration and manual
review of the following data : computer analysis using the program
Genefinder (P. Green and L. Hillier, personal communication), the
large scale EST projects of Yuji Kohara
(http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C.
elegans ORFeome cloning project (<http://wormfdb.dfc.harvard.edu/>),
similarity to other proteins from BlastX analyses
(<http://blast.wustl.edu/>), sequence conservation with C. briggsae
using Jim Kent's WABA alignment program (Genome Research
10:1115-1125, 2000), individual C. elegans GenBank submissions,
and personal communications with C. elegans researchers. tRNAs
are predicted using the program tRNAscan-SE (Lowe, T.M. and
Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

FEATURES

source

1. 45389
/organism="Caenorhabditis elegans"
/mol_type="genomic DNA"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="I"
/clone="C32E12"
750..772
/gene="mir-88"
/product="micro RNA"
/notes="standard_name=C32E12.6"
/db_xref="WormBase:C32E12.6"
972..2572
/gene="C32E12.2"

misc_RNA

gene

CDS	/locus tag="C32E12.2" join(972..1126,1184..1438,1501..1587,1633..1786, 1997..2325,2371..2572) /genes="C32E12.2" /locus tag="C32E12.2" /standard name="C32E12.2" /notes="contains similarity to Pfam domains PF01590 (GAF domain), PF0233 (3'5'-cyclic nucleotide phosphodiesterase); coded for by the following C. elegans cDNAs: OSTF082F3_1, OSTR082F3_1" /codon start=1 /product="Hypothetical protein C32E12.2" /protein_id="AA853879.1" /db_xref="GI:1703547" /db_xref="WormBase:C32E12.2" /translations="MSKGIAGYVASTGEGLIENAYEDERFNADVDSKTYTKTTLIC MPILIRIGVIVQVMNKHGDFVTRQEDAFEIFAVYCGIALHAKLYDKIRSEQKY RVALEVLAVHSVGNADENVKKEIKENIRNRIVELETIDFNGRLSELEKPLVAYMFKT LEADTLRFDTEDLIRFVLTVRKNYRVAVYHNAHGSVAHAMEFATLWNSPDFTKLEA LALVSLCHDLDRHGKNAYKMTSTPLASIYSTSVMERHFNQVTIILQDGRNAIL KLSSEYDKITLSLKHCLITDLALFFSNKALNVLINDFDINRQHRLLTQVM MTCGDLVASAKPWNIOETVKVIFEFYDQGDALRSKGKPEIPMMDRQQAHLMPQMA S"	QY	29	TGTTAGGATTTTGTATCGCACCATTTATCGGCCCAATCTCCATCGACTTCGGATCGCTCCG 88
		Db	6077	TATTTAGAAAAATTCAGCACCATTTATCGGCCCAATCTCCATCGACTTCGGATCGCTCCG 6018
		QY	89	GAGCTTTGTTGTCATCTCTCGTAGGTAAAGAGCATCAAAAACTTACCACCTGGCTCCATCAA 148
		Db	6017	GAGCTTTGTTGTCATCTCTCGTAGGTAAAGAGCATCAAAAACTTACCACCTGGCTCCATCAA 5958
		QY	149	TGGAAGCTCTTGAACGTAGTGGGTGTTCAATTTGTTGATGCTCTCATCAAAAAAGGTCAA 208
		Db	5957	TGGAAGCTCTTGAACGTAGTGGGTGTTCAATTTGTTGATGCTCTCATCAAAAAAGGTCAA 5898
		QY	209	TGGAATGCAAAAAGAGCATTTAGACCAATTAGAAGTTCTAGAGAAGTACATCCCTG 268
		Db	5897	TGGAATGCAAAAAGAGCATTTAGACCAATTAGAAGTTCTAGAGAAGTACATCCCTG 5838
		QY	269	ATCAATTCGATTAAGTACAAAAAGCTTAAAGTTGATGATTTGGCAGCTGATCGAGTTATGC 328
		Db	5837	ATCAATTCGATTAAGTACAAAAAGCTTAAAGTTGATGATTTGGCAGCTGATCGAGTTATGC 5778
		QY	329	AACAGCGGAGATGGCAAAATTAACGCCTAAA----- 360
		Db	5777	AACAGCGGAGATGGCAAAATTAACGCCTAAAATTAACGCCTAAAATTAACGCCTAAAATTA 5718
		QY	361	-----TCAGGAATGCAATTTATCGATATGTTGAATGCAAT 396
		Db	5717	CTGTTCTGAAAAAATAATTTGAATTTTCAGGAATGCAATTTATCGATATGTTGAATGCAAT 5658
		QY	397	GGATCCCAATTTGGTAGCAGTATTCGTGATAGAGATGCTATCCGAACGCGAGAGAT 456
		Db	5657	GGATCCCAATTTGGTAGCAGTATTCGTGATAGAGATGCTATCCGAACGCGAGAGAT 5598
		QY	457	ATGAAAAATACCGATCCGTCGGAAC----- 481
		Db	5597	ATGAAAAATACCGATCCGTCGGAACAGGTTGGTGTGAATTCATTTGCAGTTTTTAAAAAAT 5538
		QY	482	----- 481
		Db	5537	ATTAGAGAAATCCCTTTTCCCAATTTGTTGGTGTGAATTTATTTGTTGCAGAAATGATT 5478
		QY	482	-----AGATTGCCAAAGCCGTAATGGCAAA 507
		Db	5477	TTCTTAATTTTATACTAACTAATGTTTCTAATTTAGATTGCCAAAGCCGTAATGGCAAA 5418
		QY	508	TTTCAAAACAAATTTCTCCAGCATCTGTTGCAAAATATGATCGCTGGCAAGAACCCCTTT 567
		Db	5417	TTTCAAAACAAATTTCTCCAGCATCTGTTGCAAAATATGATCGCTGGCAAGAACCCCTTT 5358
		QY	568	AAAAATCCCTCAACAAATGAGAAAGCTCAAGCTGCTCCATCGTCAGTGTTCCAACAAGCT 627
		Db	5357	AAAAATCCCTCAACAAATGAGAAAGCTCAAGCTGCTCCATCGTCAGTGTTCCAACAAGCT 5298
		QY	628	CTTGCTCAAAAGAGCAATTTAGGTAAAAATGCCCAAGTTGCCCGTGGAAAGAGGTGAAGAA 687
		Db	5297	CTTGCTCAAAAGAGCAATTTAGGTAAAAATGCCCAAGTTGCCCGTGGAAAGAGGTGAAGAA 5238
		QY	688	CAACGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 747
		Db	5237	CAACGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5178
		QY	748	GATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 807
		Db	5177	GATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5118
		QY	808	GAAACCAAGTGAAGCAGAGCATCAGAGAAGATTTAGCCAGGAGATTTGAAAGTAGT 867
		Db	5117	GAAACCAAGTGAAGCAGAGCATCAGAGAAGATTTAGCCAGGAGATTTGAAAGTAGT 5058
		QY	868	CTTAGATTTAAAGAGCTTTTA----- 888
		Db	5057	CCTAGATTTAAAGAGCTTTTTCACAGTGAATTTTCTCAACTGTTTAAATTTTTT 4998
gene	complement(3240..6294) /genes="osr-1" /locus tag="C32E12.3" complement(join(3240..3454,3601..3712,3759..4390, 4446..4527,5034..5441,5571..5689,5742..6059,6249..6294)) /genes="osr-1" /locus tag="C32E12.3" /standard name="C32E12.3" /notes="contains similarity to Schizosaccharomyces pombe Hypothetical protein C29E6.10c in chromosome 1; SW:YAF_A SCHPO; coded for by the following C. elegans cDNAs: YK563c9.3, AY360470" /codon start=1 /product="Osmotic stress resistant protein 1" /protein_id="AM15555.1" /db_xref="GI:20198782"			
CDS	/translations="MTLFLFLLGLFCIAPLSAQSPSTSDAPGALLSLVKGSHQKL PLAPMALBMLGVQFDALIKKGOMEMAKFAKTOLEVKVHPQPFQYKKLKVD LAADVMQAEAKLOPKSGNAFIDMLNGNGIPIGSSIRGLEDAIRTORMENTDPE QIAKAMDKFTQILPGLVANMAGKPNPQPMQKQAAPSSVFQQAALAPAMLGK NAPVAGGEQEMNMRVQRMQRQLQEDDDDDLEDDVPRRSSDGEQPSAE HQRDLARRLSKPRLEQLQNAEVSLLSYQMRSDPLSKRPLANNDEDESAFRM EARKLDQKSLVLGLFGESDDDEDENIDPSENSFREAPLRLSLSGFVEKLS NDELKALDIRIKYRDVDEKYLAPKPMFNPQPGYFAPRKIPTRPKMLPLLIIGD PKQEEIRRHPSIEWKIAKESRVLTNLKNPNSLAALFMDDKLENTLKGRLMTDEQNG RTVKTRIALPRLPAPTAKEMIDAKVFODIEERPIPLFFEPKGRHTRLRWTGANE KEIPGLSRTILSLDPTMPALNTAFSTQGRARDEWDWTFKIPNNNPGDVEGFRMNS KTRFVGGNGAFDMPALG"			
gene	complement(8197..11009) /genes="sox-1" /locus tag="C32E12.5" complement(join(8197..8424,9119..9232,9281..9688, 10208..10471,10809..11009)) /genes="sox-1" /locus tag="C32E12.5" /standard name="C32E12.5" /notes="contains similarity to pfam domain PF00505 (HMG (high mobility group) box); coded for by the following C. elegans cDNAs: YK397g12.5, YK404e6.5, YK657g12.5, YK1577b07.5, YK1661e08.5, YK397g12.3, YK404e6.3, YK657g12.3, YK1577b07.3, YK1661e08.3" /codon start=1 /product="Sox (mammalian sry box) family protein 1" /protein_id="AM15556.1" /db_xref="GI:20198784"			
CDS				
Query Match .	46.3%; Score 894; DB 3; Length 45389;			
Best Local Similarity	66.7%; Pred. No. 1.9e-195;			
Matches 1894; Conservative	0; Mismatches 10; Indels 934; Gaps 6;			

Qy	889	-----	888
Db	4997	TAGAAAGGAAAAAAATCCAAAATATTTAAGCTTTATATATAAAGTATTAAGTCTTTTAT	4938
Qy	889	-----	888
Db	4937	ATGTAATTTAGAGCATATCTGTAATATAATTTTACAAATACAAATCATACATAGTGT	4878
Qy	889	-----	888
Db	4877	TGAGATCGCGGATATTGATAAATCTTCAACAGTCGGTAATACTTCTCGACGAATTATCAA	4818
Qy	889	-----	888
Db	4817	AAAAATGCCGAATCTTACTATTTTAAAGTCTATATATTTTTCGGAACCTTCTAAAAAAGT	4758
Qy	889	-----	888
Db	4757	TTTTTGTCATTTTCAAAATTTAATAATAATAGGATGCGATTTAGCTAAAGTCTCCTATTTT	4698
Qy	889	-----	888
Db	4697	GTTTGTAAGATGTTAAACCACTAATAATATTTTCGAAAAATTTTGTGGAGATCGCGAAT	4638
Qy	889	-----	888
Db	4637	TGAAAAAAATCACGTGTATCTTGGAAAAATAACTAACAAATGACAGTCGTAAAGCTAGTAG	4578
Qy	889	-----	901
Db	4577	GACATTTTGTGCTGTGTGTGCTGCTTTGAAAAAATACCTAATACTTTTCAGAAATGCGGAAG	4518
Qy	902	TTCAATCATTTGCTCTCTTTACCAACGAATGAGGGATTTCTCCACTGAGCAAGCGAAGGCTT	961
Db	4517	TTCAATCATTTGCTCTCTTTACCAACGAATGAGGGATTTCTCCACTGAGCAAGCGAAGGCTT	4458
Qy	962	TGGCTATGAAC-----	972
Db	4457	TGGCTATGAACGGTTGGTTTTTTTGTGTTTTCGAAATTAGTAGGAATCTTATCTGTGACTAA	4398
Qy	973	-----GATCAGGATGAAGTGCATTTCCGCGCAATGAGGCTCGTGCMAAATCTAGATCAA	1026
Db	4397	ATTTTCAGATGAGGATGAAGTGCATTTCCGCGCAATGAGGCTCGTGCMAAATCTAGATCAA	4338
Qy	1027	AAATCTCAACTTTGTGCTCGCTCTCCATGTTTTTGGAGAGTCTGATGATGATGAAGACGAA	1086
Db	4337	AAATCTCAACTTTGTGCTCGCTCTCCATGTTTTTGGAGAGTCTGATGATGATGAAGACGAA	4278
Qy	1087	GAGATGAAAAATTTGATGTGATCCATCTGAAAAATTCATTCGTCGTGCACCACTTCGTCCT	1146
Db	4277	GAGATGAAAAATTTGATGTGATCCATCTGAAAAATTCATTCGTCGTGCACCACTTCGTCCT	4218
Qy	1147	TCCTCCGGATTCGTTGAGAAATTTAAAGTCAAATGATGAATTTGAAAAGTGCATTTGGACAGA	1206
Db	4217	TCCTCCGGATTCGTTGAGAAATTTAAAGTCAAATGATGAATTTGAAAAGTGCATTTGGACAGA	4158
Qy	1207	ATTTAAATATCGAGTTCGTTGAAAAAGTATCTTGTCTTCCAAAGCGATGGAATTTCAAT	1266
Db	4157	ATTTAAATATCGAGTTCGTTGAAAAAGTATCTTGTCTTCCAAAGCGATGGAATTTCAAT	4098
Qy	1267	CCAAAACCTCAGGCTCGCTACTTTGTCTCCACGTTAAATCCCAACAGACCCACGTAANAATG	1326
Db	4097	CCAAAACCTCAGGCTCGCTACTTTGTCTCCACGTTAAATCCCAACAGACCCACGTAANAATG	4038
Qy	1327	CTTCCATTATTAATTTGGATCTGATCCAAAAGTTCAAGAGAAATACGAGACATCCAAGT	1386
Db	4037	CTTCCATTATTAATTTGGATCTGATCCAAAAGTTCAAGAGAAATACGAGACATCCAAGT	3978
Qy	1387	ACCGAATGGAAAAATTCGAAAAAGATCAAGAGTTTTTGCAAAATTTTGAAGATATATCCAAGT	1446
Db	3977	ACCGAATGGAAAAATTCGAAAAAGATCAAGAGTTTTTGCAAAATTTTGAAGATATATCCAAGT	3918
Qy	1447	CTTGTCTGCAATTTTCATGGATGATAAATTAGAGAAATACATTTGAAAGGAGGCAATTTGTTA	1506

Db	3917	CTTGCTGCATGTTGTCATGGATGATAAATAGAGAAATACATTGAAGAGGAAGGCAAAATGTTTA	3858
Qy	1507	ACTGATGAACAGAAAGGTAGAACACGCTGTCAAAACAAATTCGTGCATTTACCAAGACTGTTTC	1566
Db	3957	ACTGATGAACAGAAAGGTAGAACACGCTGTCAAAACAAATTCGTGCATTTACCAAGACTGTTTC	3798
Qy	1567	GGTGCAACCAACTGCAAAAGCTGAAATGATGTGACAA-----	1603
Db	3797	GGTGCAACCAACTGCAAAAGCTGAAATGATGTGACAAAGGTATAATAGAGATTAATTTATTT	3738
Qy	1604	-----AGGTATTCCAAGATATTGAAGACGTCCTCAATTCCTCC	1640
Db	3737	TTGGCAATTTAAATATTTATTTTCAGGTATTTCCAAGATATTGAAGACGTCCTCAATTCCTCC	3678
Qy	1641	ATTGTTCTTTTGAACCAAAAGGAAGGCATACGAGATTGAGATGGACTGGAGCAAAATCAAAA	1700
Db	3677	ATTGTTCTTTTGAACCAAAAGGAAGGCATACGAGATTGAGATGGACTGGAGCAAAATCAAAA	3618
Qy	1701	AGAAATTCAGGACTTT-----	1716
Db	3617	AGAAATTCAGGACTTTGGTAGGTATTGATTAAATGTTTTGTGCTTGGTTTTTCTGCAAT	3558
Qy	1717	-----	1716
Db	3557	GTCAGTGAAACATAGAAATGTGTACGAAGAACTATTGTTACGTGCGCAACACAGAAATGTA	3498
Qy	1717	-----GGAAGTCGCTTCATTCTC	1734
Db	3497	TGAGTCAAACAGAGAGAAAAACGCCAATCCTTTTCTTATAGGAAGTCGCTTCATTCTC	3438
Qy	1735	CCATCTCTTGATCCAACTATGCCAGCGTTTGAAACACGCGTTTCTCGACTCAGGGCGAGCC	1794
Db	3437	CCATCTCTTGATCCAACTATGCCAGCGTTTGAAACACGCGTTTCTCGACTCAGGGCGAGCC	3378
Qy	1795	CGTGACGATGGGATACCAATGTTCAAATCCCGAATAACTGGNAATCCTCGAGATGAAGTT	1854
Db	3377	CGTGACGATGGGATACCAATGTTCAAATCCCGAATAACTGGNAATCCTCGAGATGAAGTT	3318
Qy	1855	GGGTTCAAATGAACCTCAAAAACCAACCGATTCTGTTGGAGGAAATGGAGCATTTGATATG	1914
Db	3317	GGGTTCAAATGAACCTCAAAAACCAACCGATTCTGTTGGAGGAAATGGAGCATTTGATATG	3258
Qy	1915	CTTGCACTGGGATGTGAG	1932
Db	3257	CTTGCACTGGGATGTGAG	3240

RESULT 3

CBRG03012/c	17418 bp	DNA	linear	INV 04-NOV-2000
LOCUS	CBRG03012	17418 bp	DNA	linear
DEFINITION	Caenorhabditis briggsae cosmid G03012, complete sequence.			
ACCESSION	AC084485			
VERSION	AC084485.1	GI:11094935		
KEYWORDS	HTG.			
SOURCE	Caenorhabditis briggsae			
ORGANISM	Caenorhabditis briggsae			
	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
	Rhabditidea; Rhabditidae; Peloderinae; Caenorhabditis.			
REFERENCE	1 (bases 1 to 17418)			
AUTHORS	Washington University Genome Sequencing Center.			
TITLE	The C. briggsae Genome Sequencing Project			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 17418)			
AUTHORS	Wilson, R.			
TITLE	The sequence of C. briggsae cosmid G03012			
JOURNAL	Unpublished			
REFERENCE	3 (bases 1 to 17418)			
AUTHORS	Waterston, R.			
TITLE	Direct Submission			
JOURNAL	Submitted (04-NOV-2000) Department of Genetics, Washington			
	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA			
COMMENT	Submitted by:			

Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA
e-mail: jspieth@watson.wustl.edu

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

FEATURES

source
Location/Qualifiers
1. .17418
/organism="Caenorhabditis briggsae"
/mol_type="genomic DNA"
/strain="GujAraT G16"
/db_xref="taxon:6238"
/clone="G03012"

ORIGIN

Query Match 22.1%; Score 426.4; DB 3; Length 17418;
Best Local Similarity 80.2%; Pred. No. 1.5e-87;
Matches 515; Conservative 0; Mismatches 121; Indels 6; Gaps 1;
Qy 973 GATGAGGATCAAGTGCATTCGCGCAATCGAGCTCGTGCAAACTAGATCAAAAATCT 1032
Db 14182 GATGAACGAAACATATTCGTCGATGGAACCGCGTGCAAGCTCGATCAAAAATCG 14123
Qy 1033 CAACCTTGCTCGCTCCATGCTTTG-----GAGAGTCTGATGATGAAGACGAA 1086
Db 14122 CAATTAGTTCGGATTCCATGATTCGGAATCGGAGATGAAGATGATGACGTAG 14063
Qy 1087 GAGATGAAATTTGATGATGATCAATCTGAAATTCATTCGTCGTGCAACCACTTCGCTT 1146
Db 14062 GAGATGAGAATTCATTTGATTCATCAACCACTTCATTCGTCGTGCTCCACTTCGATTG 14003
Qy 1147 TCTTCGGATTCGTTGAGAAATTAAGTCAAAATGATGAATTTGAAAGTGCATTGGACAGA 1206
Db 14002 TCTTCGGATTCGTTGAAATTAAGTCAAAATGATGAATTTGAAAGTGCATTGGACAGA 13943
Qy 1207 ATTAATATCATGATGATGATGAGTGAAGTATCTTCTCCAAAGCGGATGGAATTCAT 1266
Db 13942 ATCAAGTATCGAGTTGACGATGTGAAAGATTTTTCGCAACCAAGCCCAATGGATTTCAAT 13883
Qy 1267 CCAAACTCAGCTGCTACTTTCCTCAGTAAATTCCTCAACCAAGCACGCTAAATG 1326
Db 13882 CCTAAACCCAGCTGCTGATATTTGTACTCGAATAATCCCACTAGACCTCGTAAATG 13823
Qy 1327 CTTCATATTAATTTGATCTGATCCAAAGTTCAAGAGAAATACCAAGACATCCAAAT 1386
Db 13822 CTTCGGTTGATCGATCAGATCCGCGACTCAAGAGATGATCCAGC 13763
Qy 1387 ACCGAATGGAATTTGCAAAAGATCAAGATTTTGAATAATTTGAAGATTAATCCAAAT 1446
Db 13762 AAAGAATGGAAGATTGCAAAAGATCAAGATTTCTGACAAATTTGAAGAACATCCAAAT 13703
Qy 1447 CTTCCTCATTTGATGATGATTAATTTAGAGATACATTTGAAGAGGCAATGTTA 1506
Db 13702 CTTCCTCATTTTATTTGATGAGAAATTTGAAATACGTTTAAGAGAGGCAATTTGTTG 13643
Qy 1507 ACTGATCAACAGAAAGTGAACACGTTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1566
Db 13642 ACTGATGAGCAAGAGGAGAACAGAGTTAAGCAATCAGATTCGTCGTCGTCGTCGTC 15383
Qy 1567 GGTGCACCAACTGCAAAAGCTGAAATGATGATGCAAAAGTA 1608
Db 13582 GGACACCAACTGCAAAAGCGGAGATGATGATGCAAAAGTA 13541

RESULT 4
CBRG33D04/c
LOCUS
DEFINITION
ACCESSION

CBRG33D04
Caenorhabditis briggsae cosmid G33D04, complete sequence.
AC084550

VERSION
KEYWORDS
SOURCE
ORGANISM

AC084550.1 GI:11095000
HTG.
Caenorhabditis briggsae
Caenorhabditis briggsae
Rhabditoidea; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 39089)
Washington University Genome Sequencing Center.
The C. briggsae Genome Sequencing Project
Unpublished
2 (bases 1 to 39089)
Wilson, R.
The sequence of C. briggsae cosmid G33D04
Unpublished
3 (bases 1 to 39089)
Waterston, R.
Direct Submission
Submitted (04-NOV-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63109, USA
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA
e-mail: jspieth@watson.wustl.edu

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

FEATURES
source
Location/Qualifiers

1. .39089
/organism="Caenorhabditis briggsae"
/mol_type="genomic DNA"
/strain="GujAraT G16"
/db_xref="taxon:6238"
/clone="G33D04"

ORIGIN

Query Match 10.9%; Score 211; DB 3; Length 39089;
Best Local Similarity 77.3%; Pred. No. 7.5e-38;
Matches 256; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
Qy 40 TGTATCGCACCATTATCGGCCCAATCTCCATCGACTTCGATCTCGGGAGCTTTGTCG 99
Db 1693 TTCTAGATCCAAATATCGCGGCTCACCACCAATTCGATGTTCCCGTGTCTTTA 1634
Qy 100 TCATCTCTCGTAGGTAAAGCCATCAAAATACCACTGGCTCCATCAATGGAAGCTTT 159
Db 1633 TCCTCTCTGCGCACTAAACCAACCCCAAACTCCATAGCTCCATCCATGGAAGCACTA 1574
Qy 160 GAATCATGGGTTCCTCAATTTGTTGATGCTCTCATCAAAAAGGTCAATGGAATGGCA 219
Db 1573 GAATTCATGGGAGTTCCTCAATTTGTTGATGCTTTAAATTAATAAGAGCAGATGGAATGGCA 1514
Qy 220 AAGGAGCATTTAAGACTCAATTAGAGTCTTAGAGAAAGTACATCTGATCAATTCGAT 279
Db 1513 AAGGAGCATTTCAAAACACAACTGGAGGTTCTGAGAAAGTACATCCGATCAATTTGAG 1454
Qy 280 AAGTACAAAAGCTAAAGTTGATGTTGGCAGCTGATGCGAGTATGCAACAGGCGGAG 339
Db 1453 AATATAGAACTGAAGTGGAGGATTTAGCGCAGATGCACTTATGCAACAGGCGAGAA 1394
Qy 340 ATGCGCAAAATACAGCTTAATCAGGAAATG 370
Db 1393 ATGTCCAAGTTGCAACCGAAACTGCTGAAG 1363

RESULT 5
166494/c
LOCUS
DEFINITION
ACCESSION

166494
Sequence 14 from patent US 5670367.
166494
linear
PAT 28-DEC-1997

ORIGIN

Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
 Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Glende, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R.,
 Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C.,
 Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,
 McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T.,
 Nlman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Norbu, C.,
 Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,
 Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
 Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S.,
 Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE JOURNAL

Submitted (18-FEB-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 161964)

REFERENCE AUTHORS

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
 Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,
 Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y.,
 Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K.,
 Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
 Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S.,
 Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
 Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
 Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., Maclean, C.,
 Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
 Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,
 Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
 Rachupka, A., Ramasamy, J., Raymond, C., Retta, R., Rise, C., Rogov, P.,
 Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C.,
 Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
 Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,
 Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
 Submitted (23-MAR-2004) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 23, 2004 this sequence version replaced gi:44886783.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L20154
 Center clone name: 502_H_14

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 9 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * been provided by the submitter.

* This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

* 1 6498: contig of 6498 bp in length
 * 6499 6598: gap of 100 bp
 * 6599 14424: contig of 7826 bp in length
 * 14425 14524: gap of 100 bp
 * 14525 35310: contig of 20786 bp in length

* 35311 35410: gap of 100 bp
 * 35411 58107: contig of 22897 bp in length
 * 58108 74207: gap of 100 bp
 * 74208 74399: contig of 16092 bp in length
 * 74399 74399: gap of 100 bp
 * 74400 103424: contig of 29025 bp in length
 * 103425 103524: gap of 100 bp
 * 103525 156902: contig of 53378 bp in length
 * 156903 157002: gap of 100 bp
 * 157003 159646: contig of 2644 bp in length
 * 159647 159746: gap of 100 bp
 * 159747 161964: contig of 2218 bp in length.

FEATURES source

Location/Qualifiers
 1. .161964
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="10"
 /map="10"
 /clone="RP24-502H14"
 /clone_lib="RPCI-24 Male Mouse BAC"

ORIGIN

Query Match 2.9%; Score 56.2; DB 2; Length 161964;
 Best Local Similarity 49.8%; Pred. No. 0.037;
 Matches 142; Conservative 0; Mismatches 143; Indels 0; Gaps 0;
 QY 635 AAAGACCAATGTTAGTAAATAATGCCCGTGGAGAGGTGAAGAACACCGA 694
 Db 54954 AAAAAAAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAGG 55013
 QY 695 TGATGATGATCGAGTGGACCAAGATGCACAAAGAGAACTTCAAGAGGAGATGAAG 754
 Db 55014 AAGAGAAAG 55073
 QY 755 ATGATGATGATCTTGTAGGACGAGATGTACCCAGAAAGAAAGTCCGATGGAGAACCCAC 814
 Db 55074 AGAAG 55133
 QY 815 AAAGTGAAGCAGCATCAGAGAAAGAGATTTAGCCAGGAGATTTGAAAGATGTCCTAGAT 874
 Db 55134 AGGAAGAGAAAG 55193
 QY 875 TAAAGAGCTTTTACAGAAATGCCGAAGTTCATCATTCCTCTT 919
 Db 55194 GAAGGATTTTTTAAAAAGTAGCACTATATATATATATCATTT 55238

RESULT 8 AC147563/c

LOCUS AC147563 236428 bp DNA linear ROD 29-MAY-2004
 DEFINITION Mus musculus BAC clone RP23-231H21 from chromosome 10, complete
 sequence.
 ACCESSION AC147563
 VERSION AC147563.4 GI:47777619
 KEYWORDS HTG.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 236428)
 AUTHORS Levy, A., Kozlowicz, A. and Meyer, R.
 TITLE The sequence of Mus musculus BAC clone RP23-231H21
 JOURNAL Unpublished (2001)
 REFERENCE 2 (bases 1 to 236428)
 AUTHORS Wilson, R.K.
 TITLE Direct Submission
 JOURNAL Submitted (17-DEC-2003) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA
 REFERENCE 3 (bases 1 to 236428)
 AUTHORS Wilson, R.K.
 TITLE Direct Submission
 JOURNAL Submitted (12-MAY-2004) Genome Sequencing Center, 4444 Forest Park

REFERENCE AUTHORS TITLE JOURNAL

Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 236428)
Wilson R.K.
Direct Submission
Submitted (28-MAY-2004) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
5 (bases 1 to 236428)
Wilson R.K.

REFERENCE AUTHORS TITLE JOURNAL

Direct Submission
Submitted (29-MAY-2004) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On May 28, 2004 this sequence version replaced gi:47131324.

COMMENT

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu>
Contact: submissions@watson.wustl.edu
----- Summary Statistics

Center project name: M_BA0231H21

NOTICE:

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
<http://genome.wustl.edu>

SOURCE INFORMATION:

The RPCI-23 BAC Library has been constructed by Kazutoyo Oseawa
and Minako Tateno in the laboratory of Pieter de Jong
(<http://www.chori.org>) from female C57BL/6J mouse kidney and/or
brain genomic DNA. The clone and detailed information can be
obtained from Research Genetics, Inc. (<http://www.resgen.com>) or
Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

FEATURES source

Location/Qualifiers
1..236428
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="10"
/map="10"
/clone="RP23-231H21"
/clone_lib="RPCI-23"

repeat_region
1..171
/rpt_family="MaLR"
repeat_region
1561..1790
/rpt_family="B4"
repeat_region
2037..2200
/rpt_family="B4"
repeat_region
3815..3958
/rpt_family="Alu"
repeat_region
4188..4637
/rpt_family="B4"
repeat_region
4635..5812
/rpt_family="B4"
repeat_region
5813..5929
/rpt_family="Alu"
repeat_region
5986..6178
/rpt_family="B4"

repeat_region
6175..6492
/rpt_family="L1"
repeat_region
7740..7813
/rpt_family="L1"
repeat_region
8115..8183
/rpt_family="Alu"
repeat_region
9381..9515
/rpt_family="Alu"
repeat_region
11530..12350
/rpt_family="L1"
repeat_region
12351..12497
/rpt_family="Alu"
repeat_region
12498..12754
/rpt_family="L1"
repeat_region
12755..13176
/rpt_family="ERVK"
repeat_region
13177..13360
/rpt_family="ERVK"
repeat_region
13361..13734
/rpt_family="ERVK"
repeat_region
13740..13974
/rpt_family="ERVK"
repeat_region
13951..13986
/rpt_family="ERVK"
repeat_region
13975..14058
/rpt_family="ERVK"
repeat_region
14151..14443
/rpt_family="ERVK"
repeat_region
14467..14864
/rpt_family="ERVK"
repeat_region
14865..15350
/rpt_family="L1"
misc_feature
21401..22231
/note="CPG island (%GC=63.5, o/e=0.75, #CPGs=55)"
repeat_region
22352..24002
/rpt_family="L1"
repeat_region
24099..24536
/rpt_family="L1"
repeat_region
24537..24935
/rpt_family="MaLR"
repeat_region
24936..25626
/rpt_family="L1"
repeat_region
25972..26123
/rpt_family="MER1_type"
unsure
26141..26511
/note="Unresolved simple sequence repeat."
repeat_region
26515..28698
/rpt_family="B4"
repeat_region
27153..27395
/rpt_family="L2"
repeat_region
29272..29412
/rpt_family="Alu"
repeat_region
30330..30671
/rpt_family="MaLR"
repeat_region
31294..31651
/rpt_family="MaLR"
repeat_region
31930..32071
/rpt_family="Alu"
repeat_region
34100..34153
/rpt_family="L1"
repeat_region
34263..34997
/rpt_family="L1"
repeat_region
34990..35758
/rpt_family="L1"
repeat_region
36065..36203
/rpt_family="B4"
repeat_region
37114..37261
/rpt_family="MER1_type"
repeat_region
37363..37416
/rpt_family="MER1_type"
repeat_region
37867..38886
/rpt_family="L1"
repeat_region
39025..39212

```

/rpt_family="B2"
complement(39136..39207)
/product="tRNA-Ser"
/note="Likely pseudogene (HMM Sc=31.53 / Sec struct
Sc=10.51)"
repeat_region
41703..41871
/rpt_family="L1"
repeat_region
41909..42096
/rpt_family="B2"
repeat_region
43371..43469
/rpt_family="ERVL"
repeat_region
45482..45576
/rpt_family="Alu"
repeat_region
45577..45594
/rpt_family="B4"
repeat_region
45648..45793
/rpt_family="Alu"
repeat_region
47665..47871
/rpt_family="MaLR"
repeat_region
47907..47995
/rpt_family="L1"
repeat_region
48419..48564
/rpt_family="MaLR"
repeat_region
48750..48858
/rpt_family="Alu"
repeat_region
49101..49272
/rpt_family="B2"
repeat_region
49922..50229
/rpt_family="MaLR"
repeat_region
50992..51294
/rpt_family="MaLR"

Query Match      2.9%; Score 56; DB 10; Length 236428;
Best local Similarity 56.4%; Pred. No. 0.041;
Matches 124; Conservative 0; Mismatches 95; Indels 1; Gaps 1;

Qy 663 AGTTCCGGTGGAGAGGTGAGAGACCGATGATGATCAATCGATGTCACCAAGAAT 722
Db 26366 AGGAGGAGAGGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 26308
Qy 723 GCACCAAGAGAGACTTCAAGAGGAGAGATGAAGATGATGATCTTGAGGACGAGGATGT 782
Db 26307 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 26248
Qy 783 ACCCAGAGAGAGAGAGTTCGGATGGAGAACCCACAAAGTGAAGCGAGAGCATCAGAGAGAG 842
Db 26247 AGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 26188
Qy 843 TTTAGCCAGGAGATTGAAGAGTAGTCTCTAGATTAAAGAG 882
Db 26187 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 26148

RESULT 9
AC116690 AC116690 171583 bp DNA linear ROD 15-DEC-2003
LOCUS Mus musculus chromosome 8, clone RP24-27916, complete sequence.
DEFINITION AC116690
ACCESSION AC116690
VERSION AC116690.15 GI:39841452
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 171583)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hapopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,X., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Menus,L., Mihova,T., Menga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-NOV-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 5 (bases 1 to 171583)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hapopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,X., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Menus,L., Mihova,T., Menga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-NOV-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

```

```

Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Menus,L.,
Mihova,T., Menga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (02-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 171583)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hapopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,X., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Menus,L., Mihova,T., Menga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (12-NOV-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 4 (bases 1 to 171583)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hapopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,X., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Menus,L., Mihova,T., Menga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-NOV-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 5 (bases 1 to 171583)
AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hapopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,X., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Menus,L., Mihova,T., Menga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-NOV-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

```


RESULT 12	AC123071/C
LOCUS	DEFINITION

```
ACCESSION AC123071
VERSION AC123071.3 GI:49258297
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 252384)
Tomlinson,C., Bielicki,L. and Haakenson,W.
The sequence of Mus musculus BAC clone RP23-70P8
Unpublished (2001)
2 (bases 1 to 252384)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (27-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 252384)
Wilson,R.K.
Direct Submission
Submitted (09-JUN-2004) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 252384)
Wilson,R.K.
Direct Submission
Submitted (26-JUN-2004) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
5 (bases 1 to 252384)
Wilson,R.K.
Direct Submission
Submitted (10-JUL-2004) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jun 26, 2004 this sequence version replaced gi:48475394.
----- Genome Center
Center: Washington University Genome Sequencing Center
Web code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.edu
----- Summary Statistics
Center project name: M_BA0070P08
-----

NOTICE:

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa
and Minako Tateno in the laboratory of Pieter de Jong
(http://www.chori.org) from female C57BL/6J mouse kidney and/or
brain genomic DNA. The clone and detailed information can be
obtained from Research Genetics, Inc. (http://www.resgen.com) or
Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone.
Location/Qualifiers
1. .252384
/organisms="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"

FEATURES
source
/chromosome="13"
/map="13"
/clone="RP23-70P8"
/clone_lib="RPCI-23"
428..769
/rpt_family="ERVK"
2931..3007
/rpt_family="ERVK"
3024..3307
/rpt_family="ERVK"
3299..4073
/rpt_family="ERVK"
4076..4417
/rpt_family="ERVK"
8061..8455
/rpt_family="MaLR"
10941..11107
/rpt_family="B2"
11852..12148
/rpt_family="L1"
13139..13243
/rpt_family="MIR"
13324..13373
/rpt_family="ERV1"
13351..13423
/rpt_family="ERV1"
13519..13591
/rpt_family="B4"
13592..13649
/rpt_family="ID"
13657..13792
/rpt_family="Alu"
13697..13815
/rpt_family="MaLR"
13793..13839
/rpt_family="B4"
14880..15115
/rpt_family="MaLR"
15122..15310
/rpt_family="MER1_type"
15319..15454
/rpt_family="Alu"
15559..15665
/rpt_family="MaLR"
15886..15945
/rpt_family="ERV1"
15959..16151
/rpt_family="B2"
16800..16902
/rpt_family="Alu"
18719..18859
/rpt_family="Alu"
18945..19413
/rpt_family="L1"
20321..20427
/rpt_family="Alu"
20854..21193
/rpt_family="L1"
21194..21557
/rpt_family="MaLR"
21561..21738
/rpt_family="L1"
22261..22863
/rpt_family="RMER6A"
23612..23752
/rpt_family="MIR"
23808..23935
/rpt_family="L2"
25086..25446
/rpt_family="L1"
25481..25539
/rpt_family="ERV1"
25663..25773
```


REFERENCE
AUTHORS
TITLE
JOURNAL

Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 220873)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (12-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 12, 2002 this sequence version replaced gi:21747305.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GWN
Center clone name: CH230-246N20
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 197817 bases at least Q40
Consensus quality: 201230 bases at least Q30
Consensus quality: 202948 bases at least Q20
Estimated insert size: 203330; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 220873: contig of 220873 bp in length.

FEATURES
source
1. .220873
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-246N20"
1. .1073
/notes="wgs_contig"
1. .767
/notes="clone boundary
clone end:T7
site:MboI
end sequence:RXAX82TJ"
1575. .2845
/notes="wgs_contig"
3347. .4452
/notes="wgs_contig"
complement[218194. .219030]
/notes="clone boundary
clone end:Sp6
site:MboI
end sequence:RXAX82TV"
ORIGIN
Query Match 2.88; Score 54; DB 2; Length 220873;

Best Local Similarity 51.2%; Pred.No.0.12;
Matches 126; Conservative 0; Mismatches 120; Indels 0; Gaps 0;
QY 706 CGATGGACCAAGAGTGCACAAAGAGAACTTCAAGACGAAGATGAAGATGATGAT 765
Db 79454 CAAGGTATGAACAATTTTACAATGGTAACTGGATGATGATGATGATGATGAT 79395
QY 766 CTTGAGGACGAGATGTACCCAGAGAGAAAGTTCCGATGGAGACCAACAAAGTGAAGCA 825
Db 79394 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 79335
QY 826 GAGCATCAGAGAGAGATTTAGCCAGGAGATTGAAAAGTAGTCTAGATTAAAAGAGCTT 885
Db 79334 GAAATATATATGATACATTTAAATCTGTATTTTATCAATTTGAAACATGCAAGGATAT 79275
QY 886 TTACAGAAATGCGGAAGTTCAATTCATTCCTTTACCAACGAATGAGGATTTCCACTG 945
Db 79274 TGACGAATTTCAAGAGGTATACCACTTACCAAAAGTTAGACCAAGTGAAGATAACAATTT 79215
QY 946 AGCAAG 951
Db 79214 TAAAG 79209
RESULT 14
AC097214/c
LOCUS
DEFINITION
AC097214
Rattus norvegicus clone CH230-150M11, *** SEQUENCING IN PROGRESS

ACCESSION
AC097214 GI:305211196
VERSION
HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 225493)
Muzny,D,Marie, Metzker,M, Lee, Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebegeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,X., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Iolebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulsegged,H., Lozano,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Mallory,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwankwelu,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Polindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,

Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, P., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhauser, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G. and Gibbs, R.A.

TITLE JOURNAL

REFERENCE AUTHORS

TITLE JOURNAL

REFERENCE AUTHORS

TITLE JOURNAL

COMMENT

Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 225493)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On May 10, 2003 this sequence version replaced gi:23617016.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GIOV

Center clone name: CH230-150M11

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 205588 bases at least Q40

Consensus quality: 209426 bases at least Q30

Consensus quality: 211934 bases at least Q20

Estimated insert size: 218092; sum-of-contigs estimation

Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

consists of 1 'contigs' sequence. The contigs

are represented as runs of N. The order of the pieces

is believed to be correct as given, however the sizes

of the gaps between them are based on estimates that have

provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 1 225493: contig of 225493 bp in length.

Location/Qualifiers

1. .225493

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

FEATURES source

/db_xref="taxon:10116"
/clone="CH230-150M11"
3680. 4488
/note="clone_boundary
clone_end:T7
site:EcoRI
end sequence:RWBau78TJC"
89244. .90370
/note="wgs_contig"
131230. .133655
/note="wgs_contig"
ORIGIN
Query Match 2.8%; Score 54; DB 2; Length 225493;
Best Local Similarity 51.2%; Pred. No. 0.12;
Matches 126; Conservative 0; Mismatches 120; Indels 0; Gaps 0;
QY 706 CGATGGGACCAAGAAAGTCAACAAAGAGAACTTCAAGGAGGATGAAGATGATGATGAT 765
Db 209175 CAAGGTTATGAACAATTTTACAAATGGTAACCTGGATGATGATGATGATGATGATGAT 209116
QY 766 CTTGAGGACGAGGATGTACCCAGAGAGAGAGTTCGGATGAGAGACCAACCAAGTGAAGCA 825
Db 209115 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 209056
QY 826 GAGCATCAGAGAGAGAGATTTAGCCAGGAGATTTGAAAGTAGTCTCTAGATTAAAGAGCTT 885
Db 209055 GAAATTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 208996
QY 886 TTACAGATCGGAGTTCATATGCTCTCTTACCAACGAAGTGAAGGATTTCTCCACTG 945
Db 208995 TGACGAATTTCAAGAGGTATACCACTTACCAAGTGTAGACCAAGTGAAGTGAAGTGAAGT 208936
QY 946 AGCAAG 951
Db 208935 TAAAG 208930

RESULT 15

AC113937

LOCUS

DEFINITION

AC113937

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC113937 225983 bp DNA linear HTG 10-MAR-2004
Mus musculus chromosome 6 clone RP23-21F2 map 6, *** SEQUENCING IN
PROGRESS ***, 9 ordered pieces.

AC113937 6 GI:45332531

HTG; HTGS PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 225983)

Birren, B., Nusbaum, C. and Lander, E.

Mus musculus chromosome 6, clone RP23-21F2

Unpublished

2 (bases 1 to 225983)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,

Anderson, S., Barna, N., Baetien, V., Boguslavsky, L., Boukhalter, B.,

Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, S.,

Choepei, Y., Collange, M., Collins, S., Collymore, A., Cook, A.,

Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,

Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,

Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,

Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,

Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R.,

Landers, T., Lehoczy, J., Levine, R., Liu, G., MacLean, C.,

Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,

McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T.,

Mienna, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,

Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,

Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,

Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,

Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S.,

Savery, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,

Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

Journal
Submitted (05-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

3 (bases 1 to 225983)
Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
Anderson, M., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T.,
Boguslavsky, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y.,
Collumore, A., Cook, A., Cooke, P., Corum, B., DeArellano, K.,
Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karakas, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., MacLean, C.,
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
Meldrim, J., Meneus, L., Mithova, T., Mienga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Rachupka, A., Ramaamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C.,
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,
Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

Journal
Submitted (10-MAR-2004) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 10, 2004 this sequence version replaced gi:44886384.
All repeats were identified using RepeatMasker:
Smit, A. F. A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIER
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L23232
Center clone name: 21_F_2

NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 12243: contig of 12243 bp in length
* 12244 12343: gap of 100 bp
* 12344 44783: contig of 32440 bp in length
* 44784 44883: gap of 100 bp
* 44884 60448: contig of 15565 bp in length
* 60449 60548: gap of 100 bp
* 60549 82671: contig of 22123 bp in length
* 82672 82771: gap of 100 bp
* 82772 127514: contig of 44743 bp in length
* 127515 127614: gap of 100 bp
* 127615 162894: contig of 35280 bp in length
* 162895 162994: gap of 100 bp
* 162995 185023: contig of 22029 bp in length
* 185024 185123: gap of 100 bp
* 185124 192636: contig of 7513 bp in length
* 192637 192737: gap of 100 bp
* 192737 225983: contig of 33247 bp in length.

FEATURES

Location/Qualifiers
1..225983
/organism="Mus musculus"

/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="6"
/map="6"
/clone="RP23-21P2"
/clone_lib="RPC1-23 Female Mouse BAC"

ORIGIN

Query Match 2.8%; Score 54; DB 2; Length 225983;
Best Local Similarity 38.1%; Ref. No. 0.12;
Matches 182; Conservative 0; Mismatches 295; Indels 1; Gaps 1;
QY 620 AACAACTCTTGTCTCAAGAGCAATCTTAGGTAAATAATCCCCAGTTGCGGTGGAGAG 679
Db 60237 AGAAGAAGAAACAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60296
QY 680 GTCAAGAAACAACCGGATGATGATGAATCGAGTGGACCAAAAGAAATGCAACAAAGAACTTC 739
Db 60297 AAGAACAACAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 60356
QY 740 AAGAGAAAGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 799
Db 60357 AAGAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 60415
QY 800 CGGATGGAGAACCAACAAGGTGAAGCAGAGCATTCAGAGAGAGAGATTTAGCCAGGAGATTGA 859
Db 60416 GAAGAAGAAGACGTGTTATGAAAGAAAGAAATCANNNNNNNNNNNNNNNNNNNNNNNN 60475
QY 860 AAAGTAGTCTAGATTAAAGAGCTTTTACAGAAATCGGAAGTTCAATCATCTCTCTT 919
Db 60476 NNN 60535
QY 920 ACCAACGAATGAGGAGATCTCCACTGAGCAAGCGAGGCTTTGGCTATGACGATGAGG 979
Db 60536 NNNNNNNNNNNNGAAAAAGAAAGAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 60595
QY 980 ATGAAAGTGCATTCCGCGCAATGGAGGCTCGTCAAAACTAGATCAAAATCTCAACTTG 1039
Db 60596 AGGAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 60655
QY 1040 TGCTCGGTCTCCATGGTTTTGGAGAGTCTGTGATGATGATGATGATGATGATGATGATG 1097
Db 60656 AAGAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 60713

Search completed: July 15, 2005, 13:09:15
Job time : 8512 secs

Result No.	Query Match	Score	Length	DB	ID	Description	
1	3.0	57.2	2000	8	ADA71938	Rice gene	
2	2.7	51.6	3653	2	AAX20280	Borelia	
3	2.7	51.2	3399	2	AAT05868	Chicken g	
C 4	2.6	50.4	193853	11	ACN44956	Mouse gene	
C 5	2.6	49.4	110000	2	AAV21209_01	Continuation (2 of	
6	2.5	48.2	77287	9	AAD58361	Murine tu	
7	2.5	47.8	167739	9	AAU58258	Murine tu	
8	2.5	47.8	13407	13	ABD33029	Mouse car	
C 9	2.5	47.8	180385	10	ADL13931	Abd33029	
10	2.4	47.2	3334	8	ACA28380	Osteoarth	
11	2.4	47	1686	2	AAQ87587	Prokaryot	
C 12	2.4	46.2	2427	5	AAS88862	DNA encod	
C 13	2.4	46.2	2427	5	AAS69554	DNA encod	
14	2.4	46	423	5	AAS90685	DNA encod	
C 15	2.4	45.8	17848	4	AAS45323	Chemical	
C 16	2.4	45.8	17848	6	ABK33976	Human che	
C 17	2.4	45.8	17848	6	ABD28164	DNA trans	
18	2.4	45.8	127767	13	ABD23657	Mouse car	
C 19	2.3	45.4	272	4	AAI12118	Probe #11	
C 20	2.3	45.4	272	4	ABA66298	Human fo	

XX 27-AUG-2003 (revised)
 DT 14-AUG-1996 (first entry)
 XX
 DE Chicken leucocytozoan DNA encoding immunogenic protein for vaccines.
 XX
 KW Chicken leucocytozoan; immunogen; recombinant vaccine; protection;
 XX immunisation; vaccination; ss.
 XX Leucocytozoan sp.
 OS
 XX Location/Qualifiers
 FH 1..3399
 CDS /*tag= a
 FT 1150..3218
 FT /*tag= b
 FT misc_feature
 FT /note= "fragment referred to in the claims, for use as
 FT insert in a recombinant vaccine against chicken
 FT leucocytozoan disease"
 FT
 XX JP07284392-A.
 PN
 XX 31-OCT-1995.
 XX
 XX 19-APR-1994; 94JP-00080643.
 PF
 XX 19-APR-1994; 94JP-00080643.
 PR
 XX (DORU-) DOBUTSUO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.
 PA (KITA) KITASATO KENKYUSHO SH.
 XX
 DR WPI; 1996-006311/01.
 DR P-PSDB; AAR97866.
 XX
 XX Chicken leucocytozoan immunogenic protein - used in a recombinant vaccine
 PT against chicken leucocytozoan disease.
 XX
 XX Claim 6; Page 6-9; 35pp; Japanese.
 PS
 XX AAT05868 encodes a chicken leucocytozoan immunogenic protein, this DNA or
 CC a fragment of it can be used in a recombinant vaccine to immunise against
 CC chicken leucocytozoan disease. The DNA is used in a vector and
 CC operatively linked to an expression regulatory sequence as in standard
 CC practice. (Updated on 27-AUG-2003 to correct OS field.)
 XX
 XX Sequence 3399 BP; 1577 A; 508 C; 798 G; 516 T; 0 U; 0 Other;
 SQ

Query Match 2.7%; Score 51.2; DB 2; Length 3399;
 Best Local Similarity 43.6%; Pred. No. 0.02;
 Matches 278; Conservative 0; Mismatches 358; Indels - 2; Gaps 1;

QY 483 GATTGCCAAAGCGGTAATGGACAAATTTCAACACACAAATTTCCCGAGGACTCGTTGCAAA 542
 DB 1570 GATTACAGAGTAATGTAGAGAAATCAGATGAACAGACATATTAACATATGAATA 1629

QY 543 TATGATCGCTGGCAAGACCCCTTTAAAA--TGCCTCAACAAATGAGAAAAAGCTCAAGCT 600
 DB 1630 GATAATGACATACAGAAGAGCATGAAAGAGTAACACATGAGAGAAAGAAAGAGAGTA 1689

QY 601 GCTCATCGTCAGTTTCCACAGACTCTTGTCTAAAGAGCAATGTTAGGTAAATAATGCC 660
 DB 1690 ACACATGAAGAAATAGAAAAAGAGAGCATGAAGAAAGTAATACATGAAGAGAAAAAGAA 1749

QY 661 CCAGTTCCGGTGGAGAGGTGAGAACACCGATGATGATGAATCGAGTCGACCAAGA 720
 DB 1750 GAAGTAACATGAGAAATGAGAAAGAGAGCATGAAGAGTAATACATGAAGAGAA 1809

QY 721 ATGCAACAAAGAGAACTTCAAGAGAGCAATGAAGATGATGATCTTTGAGGACGAGGAT 780
 DB 1810 AAAGAGAGGTAAACACATGAGAGAAAGAAAGAGAGCATGAAGAGTAATACATGAA 1869

QY 781 GTACCCAGAGAGAGTTCGGATGGAGAACCAACAAAGTGAAGCAGCAGCATCAGAGAGA 840

Db 1870 GAAGAAAAAGAGAGTAACACATGAAGAAATAGAAAAAGAGAGCATGAAGAGTAATA 1929

QY 841 GATTTAGCCAGGAGATTGAAAAGTAGTCTCTAGATTAAAGAGAGCTTTTACAGAAATCGGAA 900

Db 1930 CATGAGAGAGAAAGAAAGAGTAACACATGAAGAAATAGAAAAAGAGAGCATGAAGAA 1989

QY 901 GTTCAATCATTTGCTCTCTTTACCAAGCAATGAGGGATTCTCCACTGAGCAAGCGAAGGCCT 960

Db 1990 GTAATACATGAGGAGAAAGAAAGAGTAACACATGAAGAAATAGAAAAAGAGAGCAT 2049

QY 961 TTGGCTATGAACGATGAGGATGAAGTCCNTTCCCGCAANTGGAGGCTGCTGCAAAACTA 1020

Db 2050 GAAGAAAGTAATACATGAAGAAAGAAAGAGTAACACATGAAGAAATAGAAAAAGAA 2109

QY 1021 GATCAAAAATCTCAACTTTGCTCGTCTCATGTTTGGAGAGTCTTGATGATCATGAA 1080

Db 2110 GAGCATGAGAGTAATACATGAAGAAAGAAAGAGTAACACATGAAGAAATAGAAAAAGAA 2169

QY 1081 GACGAAGAGATGAAAAATTTGATTCATTCATTCCTGAAAA 1118

Db 2170 AAAGAGAGCATGAAGAGTAATACATGAAGAGAAAA 2207

RESULT 4
 ACN44956/c
 ID ACN44956 standard; DNA; 193853 BP.
 XX
 AC ACN44956;
 XX
 DT 18-NOV-2004 (first entry)
 XX
 DE Mouse genomic sequence MCG6056.
 XX
 KW Cytostatic; carcinoma; lymphoma; cancer; murine; gene; ss.
 XX
 OS Mus musculus.
 XX
 PN WO2003073826-A2.
 XX
 PD 12-SEP-2003.
 XX
 PF 28-FEB-2003; 2003WO-US006235.
 XX
 PR 01-MAR-2002; 2002US-00087192.
 XX
 XX (SAGR-) SAGRES DISCOVERY.
 XX
 PI Morris DW;
 XX
 DR WPI; 2003-328604/31.
 XX
 XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
 XX comprises a nucleotide sequence.
 PS
 XX Claim 1; SEQ ID NO 1663; Opp; English.
 XX
 XX The present invention relates to novel DNA and protein sequences which
 XX are associated with carcinomas. The sequences are useful for: (i) for
 XX screening drug candidates; (ii) for screening of bioactive agent capable
 XX of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
 XX a bioactive agent capable of modulating the activity of CAP; (iv) for
 XX evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
 XX carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
 XX carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
 XX (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
 XX determining Carcinoma Associated (CA) gene copy number. In addition, the
 XX CA genes are useful as DNA vaccines and the CAP are useful as markers of
 XX carcinoma including lymphoma. The present sequence is one such CA coding
 XX sequence. Note: This patent is an equivalent to basic patent
 XX US200218586A1, for which no sequence data was published
 XX
 XX Sequence 193853 BP; 44192 A; 37310 C; 40327 G; 53970 T; 0 U; 18054 Other;
 XX SQ

Query Match 2.6%; Score 50.4; DB 11; Length 193853;
Best Local Similarity 28.9%; Pred. No. 0.18; 302; Indels 0; Gaps 0;
Matches 123; Conservative 0; Mismatches 302; Indels 0; Gaps 0;

QY 673 GGAAGAGGTGAAGAACCAACGGATGATGATGAATCGAGTGGACCAAGAAATGCAACAAAGA 732
DB 34738 GAAAAGAAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 34679

QY 733 GAACCTTCAAGAGGAAGATGAAGATGATGATCTTTGAGGACGAGGATGTACCCAGAAGA 792
DB 34678 GAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 34619

QY 793 AGAAGTTCCGATGGAGAACACAAAGTGAAGCAGAGCATCAGAGAGAGATTTAGCCAGG 852
DB 34618 AGAAGNN 34559

QY 853 AGATTGAAAGTAGTCTCTAGATTAAAGAGCTTTACAGAAATCGGAAAGTTCAATCATG 912
DB 34558 NNN 34499

QY 913 CTCTCTTACCAAGATGAGGATTTCTCACTGAGCAAGCGAGCGCTTTGGCTATGAAC 972
DB 34498 NNN 34439

QY 973 GATGAGGATGAAGTCAATTCGCGCAATCGAGCTCGTCAAACTAGATCAAAATCT 1032
DB 34438 NNNNNNAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAA 34379

QY 1033 CAACCTTGTGCTCGTCTCCATGTTTGGAGAGTCTGATGATGAAGACGAAGAGAT 1092
DB 34378 GAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAA 34319

QY 1093 GAAAA 1097
DB 34318 GAAGA 34314

RESULT 5

AAV21209_01/c
Continuation (2 of 17) of AAV21209 from base 100001 (Methanococcus jannaschii circular
WP Sequence split into 17 fragments LOCUS AAV21209 Accession Aav21209

WP Fragment Name Begin End
WP AAV21209_00 1 110000
WP AAV21209_01 100001 210000
WP AAV21209_02 200001 310000
WP AAV21209_03 300001 410000
WP AAV21209_04 400001 510000
WP AAV21209_05 500001 610000
WP AAV21209_06 600001 710000
WP AAV21209_07 700001 810000
WP AAV21209_08 800001 910000
WP AAV21209_09 900001 1010000
WP AAV21209_10 1000001 1110000
WP AAV21209_11 1100001 1210000
WP AAV21209_12 1200001 1310000
WP AAV21209_13 1300001 1410000
WP AAV21209_14 1400001 1510000
WP AAV21209_15 1500001 1610000
WP AAV21209_16 1600001 1664976

Query Match 2.6%; Score 49.4; DB 2; Length 110000;
Best Local Similarity 49.2%; Pred. No. 0.25; 161; Indels 2; Gaps 1;
Matches 158; Conservative 0; Mismatches 161; Indels 2; Gaps 1;

QY 1160 TTGAGAAATTAAGTCAAATGATGAATTTGAAAAGTGGATTTGGACAGAAATTAATATCGAG 1219
DB 80139 TAGAAAATCTGATTTTACCTGTTTAAATATGTAATGCTTTTATACCTTAAATAAACTTGAG 80080

QY 1220 TTGATGACGTGGAAGATCTTCTCCAAAGCGGATGGAATTCATCCAAAACCTCAGC 1279
DB 80079 AGTATGATTATGCTATTAATAATTTATGATAAAATATTGAAAAATATCCAAAATCAGCTA 80020

QY 1280 CTGGCTACTTTGCTCCACGTAAAAATCCCAAGACCACGTAAAAATGCTTCCATTATTA 1339

DB 80019 TTCCCTGGCTGAAAAAGGGAATATTATATAGAGAGCAAACTTAAAAAATCATTAG 79960
QY 1340 TTGGATCTGATCCAAAGTTCAAGAGGAATACGAAGACATCCAAAGTACCGAATGAAAA 1399
DB 79959 AATGTTTGTAT--AATGCTTTAAAAATAATCCAAAGATTGCCAGTCCCTTATTGTACAA 79902

QY 1400 TTGCAAAAGATCAAGAGTTTTCACAAATTTGAAGAAATATCCAAAGTCTTGCTGCATTGT 1459
DB 79901 AGGAGAAATATTATTTTAAACTTTGGAAGATATGGAGAAGCATTTAAATGTCATAAAAAAGT 79842

QY 1460 TCATGATGATAAATTTAGAGA 1480
DB 79841 TTTTGAAGAAATATTAAGA 79821

RESULT 6

AAD58261

ID AAD58261 standard; DNA; 77287 BP.

XX AAD58261;

AC AAD58261;

XX 20-NOV-2003 (first entry)

XX Murine tumour suppressor gene, Lmt intron 3 DNA.

XX Tumour suppressor gene; Lmt; cancer; therapy; cytostatic; murine; ds.

XX Mus sp.

XX WO2003066869-A1.

XX 14-AUG-2003.

XX 07-FEB-2003; 2003WO-AU000126.

XX 07-FEB-2002; 2002AU-00000371.

XX (HALL-) HALL INST MEDICAL RES WALTER & ELIZA.

XX Cook WD, Mccaw BJ;

XX WPI; 2003-646311/61.

XX New nucleic acid molecule, useful for screening a subject for the

XX presence of an aberration in a gene encoding an LMT.

XX Claim 10; Page 196-218; 373pp; English.

XX The invention relates to novel tumour suppressor gene, referred to as

XX Lmt. The invention also relates to the field of cancer therapy and cancer

XX diagnostics. The nucleic acid molecule is useful for screening a subject

XX for the presence of an aberration in a gene encoding an LMT. The present

XX sequence is murine Lmt intron 3 DNA

XX SQ Sequence 77287 BP; 20906 A; 15437 C; 15922 G; 20915 T; 0 U; 4107 Other;

Query Match 2.5%; Score 48.2; DB 9; Length 77287;

Best Local Similarity 55.8%; Pred. No. 0.45;

Matches 92; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 717 AGAATGCACAAAGAAAGAACTTCAAGGGAAGATGAAGATGATGATGATCTTCGAGGACGA 776

DB 41785 AAGGAAGGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA 41844

QY 777 GGAATGTACCCAGAAAGAAAGTTTCGGATGGAGAACCAACCAAAAGTGAAGCAGAGCATCAG 836

DB 41845 AGAAGAAAGAAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA 41904

QY 837 AAGAGATTTAGCCAGGAGATTGAAAAGTAGTCTCTAGATTAAAGA 881

DB 41905 GAAGGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA 41949

CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
 CC prokaryotic essential genes. Note: the sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 3534 BP; 1742 A; 253 C; 655 G; 884 T; 0 U; 0 Other;

Query Match 2.4%; Score 47.2; DB 8; Length 3534;
 Best Local Similarity 46.0%; Pred. No. 0.22;
 Matches 160; Conservative 0; Mismatches 188; Indels 0; Gaps 0;

QY 1174 TCAATGATGATTCGAAGTGCATTCGACAGCAATTAATATCGAGTTGATGACGTGAA 1233
 DB 1204 TCTAAGGAAGTAAATTTAAAGAGAGAGAAATAAATTCATTAAATAATAGATGGACAGA 1263
 QY 1234 AAGTATCTTGCTCCAAAGCCGATGAAATTCATCCAAACCTCAGCCTGGCTACTTTCCT 1293
 DB 1264 AGAGAAAGATACAAATCTATATGAATAGATAAGAAATATAAAGATTAAATTTAGAA 1323
 QY 1294 CCACTGTAATCCCAAGACCCAGTAAATTCCTTCCATTTATTAATTTGGATCGATCCA 1353
 DB 1324 GTTTCGAATATAACAGAAAGATAAATAATAAATAAAGAGCTTAAAGAGATGATGAATTA 1383
 QY 1354 AAGTTCAGAGGAATACGAGACATCCAGTACCGAATGGAATTCGAAATGCGAAAGATCA 1413
 DB 1384 AAGTATAAGATACATTTAGAAACACAAAGAGATATAAATAAATAATTAAGATATGTTTCA 1443
 QY 1414 AGAGTTTTCGACAAATTCGAAGATAATCCAACTCTTGCTGCAATTCCTCATGATGATAAA 1473
 DB 1444 GCGAAAAGAGAGATTTAATAAATAAATCTCTCTGTAACATGATATTTTCTTGTATATA 1503
 QY 1474 TTAGAGATACATTCGAAGAGGAGCAATGTTAACTGATGACAGAAA 1521
 DB 1504 AAGATTATATAAATAAATTAATGGAATTTTAAATGATTTAAATAAA 1551

RESULT 11
 AAQ87587
 ID AAQ87587 standard; DNA; 1686 BP.

XX AAQ87587;

XX 27-AUG-2003 (revised)
 XX 19-DEC-1995 (first entry)

XX DNA encoding Leucocytozoan protozoa structural protein epitope.

XX leucocytozoan protozoa; structural protein; epitope; vaccine; fowl;
 XX leucocytozoanosis; treatment; ss.

XX Leucocytozoan.

XX JP07089995-A.

XX 04-APR-1995.

XX 10-SEP-1993; 93JP-00226078.

XX 10-SEP-1993; 93JP-00226078.

XX (DOBU-) DOBUTSUYO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.
 XX (NISS-) NISSEIKEN KK.

XX

DR WPI; 1995-167252/22.

DR P-PSDB; AAR70491.

XX Immune inducing polypeptide against Leucocytozoan protozoa - useful in
 PT production of vaccines for treatment of leucocytozoanosis in fowl.

XX Claim 1; Page 12-14; 20pp; Japanese.

XX AAQ87587-89 encode polypeptides having a whole or partial epitope of a
 CC structural protein of Leucocytozoan protozoa (see AAR70491-93). The
 CC polypeptides and DNA encoding them are useful in the production of
 CC vaccines for the treatment of leucocytozoanosis of fowl. (Updated on 27-
 CC AUG-2003 to correct OS field.)

XX Sequence 1686 BP; 915 A; 129 C; 382 G; 260 T; 0 U; 0 Other;

Query Match 2.4%; Score 47; DB 2; Length 1686;
 Best Local Similarity 44.2%; Pred. No. 0.18;
 Matches 194; Conservative 0; Mismatches 245; Indels 0; Gaps 0;

QY 652 AAAAATCCCCAGTTCCGGTGGAGAGGTCAAGAACCAACCGATGATGATCAATCGAGTG 711
 DB 124 AAAGAAGAACAAAG 183
 QY 712 GACCAAGAATGTCAACAAAGAGAACTTCAAGAGGAAGATGAAGATGATGATCTTCTGAG 771
 DB 184 GAACAAGAACAGATGAAG 243
 QY 772 GACGAGGATGTACCCAG 831
 DB 244 GAAGAAG 303
 QY 832 CAGAGAGAGAGATTTAGCCAGGAGAGATTTGAAAAGTAGTCTCTAGATTTAAAGAGAGCTTTTACAG 891
 DB 304 GATGAAGATGAA 363
 QY 892 AATCGGAAGTTCAATCATTTGCTCTCTTACCAACGAATGAGGGATTCTCCACTCAGCAAG 951
 DB 364 GAAGAAGATGAA 423
 QY 952 CGAAGGCTTTGGCTATGAACGATGAGAGATGAAGAGTGAATTCGCCGCAATGGAGGCTCGT 1011
 DB 424 GATGAACAAATGAAGATGAA 483
 QY 1012 GCRAAACTAGATCAAAATCTCACTTGTGCTCGTCTCTCATGTTTGGAGAGCTCTGAT 1071
 DB 484 GAAGAACAGATGAAGATGAA 543
 QY 1072 GATGATGAAGACGAGAGAG 1090
 DB 544 GATGAAGACAGAGAGAG 562

RESULT 12

AA88862/c

ID AA88862 standard; cDNA; 2427 BP.

XX AA88862;

XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #24666.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX

[illegible]

```
XX
SQ Sequence 17848 BP; 5055 A; 211 C; 3533 G; 9043 T; 0 U; 6 Other;
Query Match      2.4%; Score 45.8; DB 4; Length 17848;
Best Local Similarity 48.6%; Pred. No. 1;
Matches 125; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

Qy 1340 TTGGATCTGATCCAAAAGTTCAAGAGGAATACGAGACATCCAGTACCGAATGGAAAA 1399
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 7119 TTATACTATTAAATAAATTTAAAAAATAAATAAATAAATAAATAAATAAATAAATA 7060
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1400 TTGCAAAAGAAATCAAGAGTTTTCACAAATTTGAAGAAATAATCCAAGTCTTGTGCTATTGT 1459
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 7059 TAACAAAAAATAACGATTTCAAAAAAATTTTAAAAACGAAATTAACAAATATAATAAC 7000
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1460 TCATGGATGATAATTAGAGAAATACATTTGAAAGGAGGCAAAATGTTAACTGATGAACAGA 1519
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 6999 CACTTAAAAATAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 6940
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1520 AAGGTAGAACACGTGTCAAAACAAATTCGTGCATTACCAAGACTGTTCCGTGCACCACTG 1579
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 6939 AAAATAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 6880
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1580 CAAAAGCTGAAATGATT 1596
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 6879 TAAACATATAAATAATT 6863
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: July 15, 2005, 15:15:32
Job time : 1078 secs


```
Best Local Similarity 5.4%; Pred. No. 2.4e-11;
Matches 22; Conservative 237; Mismatches 148; Indels 0; Gaps 0;

Qy 622 CAAGCTCTTGCTCAAGAGCAATGTTAGTAAATCCCCAGTTGCCGGTGAAGAGGT 681
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1466 CAAGTAGTTAAAGAGATAGAAGATTTGGTACRRRRRRRRRRRRRRRRRRRRR 1407

Qy 682 GAAGAACACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 741
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1406 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1347

Qy 742 GAGGAAGATGAAGATGATGATCTTTGAGGACGAGGATGTACCAGAACGAAGTTCG 801
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1346 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1287

Qy 802 GATGAGAACACAAAGTGAAGACAGCATCAGAGAAAGATTTAGCCAGAGATTGAAA 861
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1286 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1227

Qy 862 AGTAGTCTAGATTAAAGAGCTTTTACAGAAATGCGGAAGTTCATCATGCTCTTAC 921
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1226 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1167

Qy 922 CAACGAATGAGGATCTCCACTGAGCAAGCAAGGCGCTTTGGCTATGAACGATGAGAT 981
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1166 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1107

Qy 982 GAAAGTCATTCGCCGCAATGGAGGCTCGTGCAAACTAGATCAAAA 1028
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1106 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1060

RESULT 2
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
; TITLE OF INVENTION: jannaschii
; Patent No. 6503729
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g

; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
```



```
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
```

```
; FEATURE:  
; NAME/key: misc_feature  
; LOCATION: (1349473)..  
; OTHER INFORMATION: n equals a, t, c, or g
```

	Query Match	2.6%;	Score 49.4;	DB 4;	Length 1664976;
	Best Local Similarity	49.2%;	Pred. No. 0.064;		
	Matches 158;	Conservative 0;	Mismatches 161;	Indels 2;	Gaps 1;
QY	1160 TTGAGAAATTAAGTC	AATAATGATGAATTGAAAAGTGCAATTTGCACAGAATTAATAATCGAG	1219		
Db	180139 TAGAABAATCTCATTTACTGTGTTAAATATGTAATGCCCTTATCTAAAAAACCTGGAG	180080			
QY	1220 TTGATGACGTGAAAAGTATCTTGCTCCAAAGCCGATGGAATTTCAATCCAAACCTCAGC	1279			
Db	180079 AGTATGATTTATGCATTTAAAAAATTTATGTATAAATATTGAAAAAAAAATATCCAAATCAGCTA	180020			
QY	1280 CTGGCTACTTTTGCTCCACGTAAAAATCCCNAACAAGACCAGCTHAAATGCTTCATTATTAA	1339			
Db	180019 TTGCGCTGGCTGAAAAGGGGGAATATTATATAGAGAAGGCCAACTTAAAAAATCATTAG	179960			
QY	1340 TTGGATCTGTATCCAAAAGTTTCAAGAGGAAATACGAAGACATCCAAGTACCGAATCGAAA	1399			
Db	179959 AATGTTTTGAT--AATGCTTTAAAAATAAATCCCAAAGATTGCCAGTCTTATTGTACAA	179902			
QY	1400 TTGCAAAAAGAAATCAAGAGTTTTTGACAAAATTTTGAAGAATAATCCAAGTCTTGCTGCAATGT	1459			
Db	179901 AGGAGAAATATTATTTTAAACTTGAAGATATCGAGAGCATTTAAATGCTTAAAAAAGT	179842			
QY	1460 TCATCGATGATAAATTAGAGA	1480			
Db	179841 TTTTGAAGAATAATAAAGA	179821			

```

RESULT 4
US-09-806-708B-22
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: Consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

```

	Query Match	2.5%	Score 47.6;	DB 4;	Length 1141;
	Best Local Similarity	9.5%;	Pred. No. 0.0016;		
	Matches 36;	Conservative 172;	Mismatches 169;	Indels 0;	Gaps 0;
Qy	1145	TTTCTTCGGAGTTCGTTGAGAAATTAAGTCAAATGATGAATTCGAAAGTCATTGGACA	1204		
Db	83	YKCKWYBCANNYSRYHARRWDMKTA YBMTTWKNGTGRHRYWKRAMBDDTVDHYYV	142		
Qy	1205	GAATTAATAATPCGATTCGATCACGTCGAAAGATGATCTTGCTCCAAAGCCGATGGAATTC	1364		
Db	143	TAMNNAWTTTMCMDKDDKRTWKKNNNAATGWDDDTKYHMMNNNGCBTVMVRYKTDR	202		
Qy	1265	ATCCAAACCTTCAGCTGGCTACTTTTGCTCCACGTAAATCCCAACCAAGACCGAGTAAA	1324		
Db	203	DWSBKRMNYGNBWBKNWSYDVYTVVWMDMCKEKVRWVTRTCRGMNVYVWBTARHR	262		

Qy	1325	TGCTTCATTATTAATTGGATCTGATCCAAAAGTTCAAGAGGAAATACGAAGACATCCAA	1384
Db	263	YNNGTBMAAYRRWTMNNNNNAKCMKRAKYGWNRAVNSTCTTWSKTTTKVRTSCWA	322
Qy	1385	GTACCGAATGGAAAAATTGCAAAAGAAATCAAGAGTTTGTGACAAATTTTGAAGAAATAATCCAA	1444
Db	323	NNCAGDANKDHKKWSAAGVYNNNNNNNTYKKARHBARDWVWHSAAKKWHANAA	382
Qy	1445	GTCTTGCTGCTATGTTTCATCGATGATAAAATTAGAGAAATACATCGAAAGGAAGGCAATGT	1504
Db	383	HYSEKKWTBYKRTTVNNNGTTTWKKEWAWYWKMDMDWEGTYNNNNNGRTYYGWTGNK	442
Qy	1505	TAACGTGATGAACAGAAA	1521
Db	443	KMIITYKKWANNCKWRA	459

RESULT 5
US-09-248-796A-6002
; Sequence 6002, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 6002
; LENGTH: 1086
; TYPE: DNA
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (987), (988)
; OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unknown
US-09-248-796A-6002

Query Match	2.4%	Score 46.8;	DB 4;	Length 1086;
Best Local Similarity	50.9%;	Pred. No. 0.0028;		
Matches 111;	Conservative 0;	Mismatches 107;	Indels 0;	Gaps 0;
Qy	131	TACCACCTGGCTCCATCAATCGAAGCTCTTGAACTGATGGGTGTTCAATTTCTTGATGCTC	190	
Db	626	TACCAATGGCTTTAGCAAGTGGTGCTGATGCATTTAATGGGTATTTCGCTGGTGGTGAAT	685	
Qy	191	TCATCAAAAAGGFTCAATGGAAATGCCAAAAGGAGCATTTAAGACTCAATTAGAAGTTC	250	
Db	686	TCATGTCTCAAGGTGAAATACGGCAAGGAGAAAATTCATTGGAAGCAACTCCAGAAATGA	745	
Qy	251	TAGAGAAGTACATCTCTGATCAATTCATTAAGTACAAAAGCTAAAGTTCGATGTTGG	310	
Db	746	AACNATTAATGGAAACAAGATGATTTAGCTATAGAAGTAGATCATGCTGATTTTGAATGGG	805	
Qy	311	CAGCTGATCGAGTTATGCAACAGCGCGAGATGGCAAAA	348	
Db	806	AAACTTTTGGCGATGAAGAAGATCCAGATGAAGAAAAA	843	

RESULT 6
US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 6259558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihc

```

; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;
; TITLE OF INVENTION: FUSIONS
;
; FILE REFERENCE: 00786/350003
;
; CURRENT APPLICATION NUMBER: US/09/007,005B
;
; CURRENT FILING DATE: 1998-01-14
;
; EARLIER APPLICATION NUMBER: 60/035,963
;
; EARLIER FILING DATE: 1997-01-27
;
; EARLIER APPLICATION NUMBER: 60/064,491
;
; EARLIER FILING DATE: 1997-11-06
;
; NUMBER OF SEQ ID NOS: 33
;
; SOFTWARE: Fast-SEQ for Windows Version 4.0
;
; SEQ ID NO 17
;
; LENGTH: 289
;
; TYPE: RNA
;
; ORGANISM: Artificial Sequence
;
; FEATURE:
;
; OTHER INFORMATION: Translation template
;
; FEATURE:
;
; NAME/Key: misc feature
;
; LOCATION: (1)..(289)
;
; OTHER INFORMATION: n = A,T,C or G
;
US-09-007-005-17

```

	Query Match	2.4%;	Score 46;	DB 3;	Length 289;
	Best Local Similarity	10.0%;	Pred. No. 0.002;		
	Matches 22;	Conservative 93;	Mismatches 105;	Indels 0;	Gaps 0;
Qy	621	ACAAGCTCTTGCTCAAAGACCAATGTTAGGTAAAAATGCCCACTGCCTGGTGGAAAGG	680		
Db	67	RSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRN	126		
Qy	681	TGAAGAACAACGGTAGTGTGTAATCGAGTGGACCAAGAATGCAACAAAGAACATTCA	740		
Db	127	RSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRN	186		
Qy	741	AGAGGAAGATGAACATCATGATGATCTTGAGGACGAGGATCTCCAGAGAAGAAGTTC	800		
Db	187	RSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRN	246		
Qy	801	GGATGGAGAACACCACAAAGTGAAGCAGACGATCAGAGAAGA	840		
Db	247	RURUGRCGRGCRUAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	286		

```

RESULT 7
US-09-244-796-17
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihc
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; TITLE OF INVENTION: FUSIONS
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
; NAME/KEY: misc feature

```

```

; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match      2.4%; Score 46; DB 3; Length 289;
Best Local Similarity 10.0%; Pred. No. 0.002;
Matches 22; Conservative 93; Mismatches 105; Indels 0; Gaps 0;

Qy 621 ACAAGCTCTTCCTCAAAGAGCAATGTTAGGTAAATGCCCCAGTTCGGGTGGNAGAGG 680
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 67 KSRVNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRN 126
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 681 TGAAGAACAACGGATGATGTAATCGAGTGGACCAAGAATGCCAACAAAGAGACTTCA 740
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 127 RSRVNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRN 186
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 741 AGAGGAAGATGAACATGATGATGATCTTGAGGACGAGATGTACCCAGAAGAGATTC 800
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 187 KSRVNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRNSRNRN 246
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 801 GGATGGAGAACAACAAAGTGAAGCAGAGCATCAGAGAAGA 840
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 247 RUTURGRCBGRCTUAAAAAATAAAAAAAAAAAAAAAAAAAAAA 286
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 8
US-09-949-016-14193/c
; Sequence 14193, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14193
; LENGTH: 247781
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(247781)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14193

```

[illegible]

RESULT 9
US-09-949-016-86857/c


```
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 91475
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-91475

Query Match      2.3%; Score 43.6; DB 4; Length 601;
Best Local Similarity 53.5%; Pred. No. 0.57;
Matches 91; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 422 GTGATTAGAGATGCTATCCGAAACGAGAGATATGGAATAATCGGATCCGTCGGAAC 481
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 482 AGATTGCGAAAGCGGTAATGGACAAATTTCAACACAAATTTCCCGAGGACTGGTTGCAA 541
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 156 AGAATATCAAGAATTTGTGGACATATTTTAAACTACCATGCTATGTTTAAATTATGCGAA 215
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 542 ATATGATCGCTGCGAAGAACCCCTTTAAATGCTCAACAAATGAGAAA 591
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 216 GTATGGTCACTGTAGAAAAAATCGAANAATAGGACAGCAAAAAA 265
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13
US-09-949-016-14370
; Sequence 14370, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 14370
; LENGTH: 107980
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(107980)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14370

Query Match      2.3%; Score 43.6; DB 4; Length 107980;
Best Local Similarity 53.5%; Pred. No. 0.57;
Matches 91; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 422 GTGATTAGAGATGCTATCCGAAACGAGAGATATGGAATAATCGGATCCGTCGGAAC 481
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 7404 GGGAAATCAGTTAACCCAGCCCCCACTCAAGGGATAGGGACTGGGCTCCACCTTTTGAAG 7463
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 482 AGATTGCGAAAGCGGTAATGGACAAATTTCAACACAAATTTCCCGAGGACTGGTTGCAA 541
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 7464 AGAATATCAAGAATTTGTGGACATATTTTAACTACCATGCTATGTTTAAATTATGCGAA 7523
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 542 ATATGATCGCTGCGAAGAACCCCTTTAAATGCTCAACAAATGAGAAA 591
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 7524 GTATGGTCACTGTAGAAAAAATCGAANAATAGGACAGCAAAAAA 7573
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 14

US-09-248-796A-3448
; Sequence 3448, Application US/09248796A
; Patent No. 6747137

GENERAL INFORMATION:

; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 3448
; LENGTH: 1755
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-3448

Query Match 2.2%; Score 43.2; DB 4; Length 1755;

Best Local Similarity 58.6%; Pred. No. 0.047;

Matches 75; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 692 GGATGATGATGAATCGATCGAGTGGACCAAGAAATGCAACAAAGAGAACTTCAAGAGGAAGATG 751

Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1625 GGAGGAAGAGTAATAATAATAATCAAAATAATTAAGAAGAAGGATTTGAACATAATGATG 1684

Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 752 AGATGATGATGATCTTTGAGGACGAGGATGTACCCAGAGAGAAAGTTCCGGATGGAGAAC 811

Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1685 ACCATGATGATGATGAAGAGGATGAAGTGAACGAGAATTTCCAACCTTATGGATATC 1744

Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 812 CACAAAGT 819

Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1745 AACAGAA 1752

Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15

US-09-820-007-3/c

; Sequence 3, Application US/09820007

; Patent No. 6830900

; GENERAL INFORMATION:

; APPLICANT: YAN, Chunhua et al

; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,

; FILE REFERENCE: CL001205

; CURRENT APPLICATION NUMBER: US/09/820,007

; CURRENT FILING DATE: 2001-03-29

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: Fast-Seq for Windows Version 4.0

; SEQ ID NO 3

; LENGTH: 213456

; TYPE: DNA

; ORGANISM: Human

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)-(213456)

; OTHER INFORMATION: n = A,T,C or G

US-09-820-007-3

Query Match 2.2%; Score 42.2; DB 4; Length 213456;

Best Local Similarity 51.9%; Pred. No. 2.4;

Matches 95; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 1299 TAAATCCCAAGAACCCAGTAAATGCTTCCATTAATTTGGATCTGATCCAAAGT 1358

Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 102305 TTAATAAATAACAAATGACAGGTTTACTGTTTTATGATATACTTTGGTAGATAACAGTTT 102246

Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1359 TCAAGAGGAATACGAAGACATCCAAAGTACCGAATGAAAATTTGCAAAAGAAATCAAGAGT 1418

Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

THIS PAGE BLANK (uspto)

Result No.	Query	Score	Query			DB	ID	Description
			Match	Length	Time			
1	1932	100.0	1932	21	US-10-736-868-1	Sequence 1, Appl		
c	50.4	2.6	193853	13	US-10-087-192-1663	Sequence 1663, Ap		
3	47.2	2.4	3534	17	US-10-382-1228-18250	Sequence 18250, Ap		
c	46.2	2.4	96602	17	US-10-085-117-61	Sequence 61, Appl		
5	45.8	2.4	17848	14	US-10-239-676-28	Sequence 28, Appl		
6	45.8	2.4	17848	15	US-10-340-453-38	Sequence 38, Appl		
c	45.8	2.4	17848	17	US-10-257-166-58	Sequence 58, Appl		

Db 61 CAATCTCATCGACTTCGGATGCTCCGGAGCTTTGTTGTCACTCTCTCGTAGGTAAAGC 120
Qy 121 CATCAAAAACTACCACTGGCTCCATCAATGGAAGCTCTTGAACTGATGGGTGTTCAATTT 180
Db 121 CATCAAAAACTACCACTGGCTCCATCAATGGAAGCTCTTGAACTGATGGGTGTTCAATTT 180
Qy 181 GTTCATGCTCTCATCAAAAAAGGTCAAAATGGAATGCAAAAGGAGCATTTAAGACTCAA 240
Db 181 GTTCATGCTCTCATCAAAAAAGGTCAAAATGGAATGCAAAAGGAGCATTTAAGACTCAA 240
Qy 241 TTAGAAGTTCTAGAGAAAGTACATCTCGATCAATTCGATAAGTACAAAAAGCTAAAAAGTT 300
Db 241 TTAGAAGTTCTAGAGAAAGTACATCTCGATCAATTCGATAAGTACAAAAAGCTAAAAAGTT 300
Qy 301 GATGATTTGGCAGCTGATGCGATTATGCAACAGCGGAGATGGCAAAATTTACAGCTTAAA 360
Db 301 GATGATTTGGCAGCTGATGCGATTATGCAACAGCGGAGATGGCAAAATTTACAGCTTAAA 360
Qy 361 TCAGGAATGCATTTATCGATATGTTGAAATGGAATGGAATCCCAATTTGGTACGATTT 420
Db 361 TCAGGAATGCATTTATCGATATGTTGAAATGGAATGGAATCCCAATTTGGTACGATTT 420
Qy 421 CGTGGATTAGAAAGTCTATCCGAACGACAGAGATATGAAATACGGATCCGTCGGAA 480
Db 421 CGTGGATTAGAAAGTCTATCCGAACGACAGAGATATGAAATACGGATCCGTCGGAA 480
Qy 481 CAGATTGCGAAGCGGTAAATGGAACAAATTTCAAAACAAATTTCTCCAGGACTCGTTGCA 540
Db 481 CAGATTGCGAAGCGGTAAATGGAACAAATTTCAAAACAAATTTCTCCAGGACTCGTTGCA 540
Qy 541 AATATGATCGTGGCAGAGAACCCCTTTAAATGCTCTCAACAAATGCAAAAGCTCAAGCT 600
Db 541 AATATGATCGTGGCAGAGAACCCCTTTAAATGCTCTCAACAAATGCAAAAGCTCAAGCT 600
Qy 601 GCTCCATCGTCAAGTCTTCCAAAGCTCTTGCTCAAAAGAGCAATGTTAGGTAAAAATGCC 660
Db 601 GCTCCATCGTCAAGTCTTCCAAAGCTCTTGCTCAAAAGAGCAATGTTAGGTAAAAATGCC 660
Qy 661 CCAGTTGCGGTGGAAGAGTGAAGAACCAACGGATGATGATGATGATGATGATGATGATGAT 720
Db 661 CCAGTTGCGGTGGAAGAGTGAAGAACCAACGGATGATGATGATGATGATGATGATGATGAT 720
Qy 721 ATGCAACAAAGAGAACTTCAAGAGGAAGTGAAGATGATGATGATGATGATGATGATGATGAT 780
Db 721 ATGCAACAAAGAGAACTTCAAGAGGAAGTGAAGATGATGATGATGATGATGATGATGATGAT 780
Qy 781 GTACCCAGAGAAAGTTCGGATGGAGAACCAAAAGTGAAGCAGAGCATCAGAGAAGA 840
Db 781 GTACCCAGAGAAAGTTCGGATGGAGAACCAAAAGTGAAGCAGAGCATCAGAGAAGA 840
Qy 841 GATTAGCCAGGAGATTGAAAAGTAGTCTTAGATTAAAGAGCTTTTACAGAAATGCGGAA 900
Db 841 GATTAGCCAGGAGATTGAAAAGTAGTCTTAGATTAAAGAGCTTTTACAGAAATGCGGAA 900
Qy 901 GTTCAATCATTTGCTCTTACCAACGAATGAGGATCTCCACTGAGCAACGAGGCGCT 960
Db 901 GTTCAATCATTTGCTCTTACCAACGAATGAGGATCTCCACTGAGCAACGAGGCGCT 960
Qy 961 TTGCTATGAAACGATGAGGATGAAAGTGCATTCGCGCAATGGAAGGCTCGTGCAAAACTA 1020
Db 961 TTGCTATGAAACGATGAGGATGAAAGTGCATTCGCGCAATGGAAGGCTCGTGCAAAACTA 1020
Qy 1021 GATCAAAAATCTCAACTTGTGCTCGGTCTCATGGTTTGGAGAGTCTGATGATGAA 1080
Db 1021 GATCAAAAATCTCAACTTGTGCTCGGTCTCATGGTTTGGAGAGTCTGATGATGAA 1080
Qy 1081 GACCAAGAAGATGAAATTTGATTGATCTCAAAATTCATTCGCTCGTGCAACCACTT 1140
Db 1081 GACCAAGAAGATGAAATTTGATTGATCCATCTGAAAATTCATTCGCTCGTGCAACCACTT 1140
Qy 1141 CGTCTTTCTCCGATTCGTTGAGAAATTAAGTCAAAATGATGAAATTTGAAAAGTGCAATTG 1200
Db 1141 CGTCTTTCTCCGATTCGTTGAGAAATTAAGTCAAAATGATGAAATTTGAAAAGTGCAATTG 1200

Qy 1201 GACAGAAATTAATATCGAGTTGATGAGCTGGAAAAGTATCTTCTCCAAAGCCGATGGAA 1260
Db 1201 GACAGAAATTAATATCGAGTTGATGAGCTGGAAAAGTATCTTCTCCAAAGCCGATGGAA 1260
Qy 1261 TTCAATCCAAAACCTCAGGCTGGCTACTTTGCTCCACGTAATAATCCCAACAGACCACGT 1320
Db 1261 TTCAATCCAAAACCTCAGGCTGGCTACTTTGCTCCACGTAATAATCCCAACAGACCACGT 1320
Qy 1321 AAAATGCTTCCATTTAATTTGATCTGATCCAAAAAGTTCAAGAGGAAATACGAAGACAT 1380
Db 1321 AAAATGCTTCCATTTAATTTGATCTGATCCAAAAAGTTCAAGAGGAAATACGAAGACAT 1380
Qy 1381 CCAAGTACCGAATGGAATAATTCGAAAAGAAATCAAGAGTTTGAACAAATTTGAAGAAAT 1440
Db 1381 CCAAGTACCGAATGGAATAATTCGAAAAGAAATCAAGAGTTTGAACAAATTTGAAGAAAT 1440
Qy 1441 CCAAGTCTTGTCTGCTCATTTGTTTCATGGATGATAAATTAGAGAATAACATTTGAAAAGGAGCAA 1500
Db 1441 CCAAGTCTTGTCTGCTCATTTGTTTCATGGATGATAAATTAGAGAATAACATTTGAAAAGGAGCAA 1500
Qy 1501 ATGTTAACTGATGAAACAGAAAGGTAGAACACGTTGTCACAAACAAATTCGTGCAATTACCAAGA 1560
Db 1501 ATGTTAACTGATGAAACAGAAAGGTAGAACACGTTGTCACAAACAAATTCGTGCAATTACCAAGA 1560
Qy 1561 CTGTTGCGTGCACCAACTGCAAAAGCTGAAATGATTTGATGCAAAAGGTATTTCCAAGATATT 1620
Db 1561 CTGTTGCGTGCACCAACTGCAAAAGCTGAAATGATTTGATGCAAAAGGTATTTCCAAGATATT 1620
Qy 1621 GAAAGACGTCCTCATTTCTTTGAAACCAAAAGGAGGATACGAGATTGAGA 1680
Db 1621 GAAAGACGTCCTCATTTCTTTGAAACCAAAAGGAGGATACGAGATTGAGA 1680
Qy 1681 TGGACTGGAAGCAAAATGAAAAGAAATTCAGGACTTTGGAAGTCCGTTCCATCTTCCCCTCT 1740
Db 1681 TGGACTGGAAGCAAAATGAAAAGAAATTCAGGACTTTGGAAGTCCGTTCCATCTTCCCCTCT 1740
Qy 1741 CTTGATCCAACTATGCGCCTTTGAACAGGCTTTTCCGACTCAGGCGGAGCCGCTGAC 1800
Db 1741 CTTGATCCAACTATGCGCCTTTGAACAGGCTTTTCCGACTCAGGCGGAGCCGCTGAC 1800
Qy 1801 GAGTGGGATACCAATGTTTCAAAATCCCGAATACTCGAATCCTCGAGATGAAGTTGGGTTT 1860
Db 1801 GAGTGGGATACCAATGTTTCAAAATCCCGAATACTCGAATCCTCGAGATGAAGTTGGGTTT 1860
Qy 1861 AAAATGAATCAAAAACCAAAACGATTCGTTGGAGGAAATGGAGCATTTGATATGCTGCA 1920
Db 1861 AAAATGAATCAAAAACCAAAACGATTCGTTGGAGGAAATGGAGCATTTGATATGCTGCA 1920
Qy 1921 CTGGGATTTGAG 1932
Db 1921 CTGGGATTTGAG 1932

RESULT 2

US-10-087-192-1663/C
; Sequence 1663, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT FILING DATE: 2002-03-01
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1663

```

; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining prior application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16250
; LENGTH: 3534
; TYPE: DNA
; ORGANISM: Clostridium botulinum
US-10-282-122A-16250

Query Match      2.4%; Score 47.2; DB 17; Length 3534;
Best Local Similarity 46.0%; Pred. No. 0.64;
Matches 160; Conservative 0; Mismatches 188; Indels 0; Gaps 0

Qy 1174 TCAAAATGATGAATTGAAAAGTCGATTTGGACACGAATTTAAATATCGAGCTTCGATGACGPG
Db 1204 TCTAAGGAAGTAAATTTAAAGAGAAGAAATTAATTCATTAAAAATAGATGGAGAC
Qy 1234 AAGTATCTTGCTCCAAAGCGGATGGAATTCAAATCCAAAACCTCAGCCTGGCTACTTTT
Db 1264 AGAGAAAGATACAAAACTATATGAATAGATAAAGATATAAAGATTAAATTTA
Qy 1294 CCACGTAAATCCCAACAAGACCGTAAATGCTTTCCATTATTAATTGGATCTGAT
Db 1324 GTTTGCAATATAACAGAAAGAAATAAATATATAAAATAAAGAGCTTAAAGAGATGATGAA
Qy 1354 AAAGTTCAAGAGGAATATCGAAGACATCCAAGTACCGATGGAATATCCAAAGAA
Db 1384 AGATATAAGATACATTAGAAACACAAAGAGATATAAATAAAAAATTAGAATATGTT
Qy 1414 AGAGTTTTCACAAATTTGAAGAATATATCCAAGTCTTGCTGCAATTTGTTTCATGGATGAT
Db 1444 GCGAAAGAGAGATTATATAAAAACTCTCCTGGTAACAATGATATTTTGGCTGAT
Qy 1474 TTAGAGATACATTGAAAGGAAGGCAAAATGTTTAACATGATGAACAGAAA 1521
Db 1504 AAAGATTATATAAATAAATTAATGGAATTTTTTAATGATTTTAAATAAAA 1551

RESULT 4
US-10-085-117-61/c
; Sequence 61, Application US/10085117
; Publication No. US2003023234A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: 529452000121
; CURRENT APPLICATION NUMBER: US/10/085,117
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 61
; LENGTH: 96602
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: variation

```


Db	530	ATGGAAGACTGTATTAGACAAACTTAAGAGAAATTCCTCATGATCAAGTACAAAGAGATT	589
Qy	288	AAAGCTAAAAGTTGATGATTTGGCAGCTGATGCAG	322
Db	590	AAAAATAAGCTATGATGGTTTAAAGTGATGATACAG	624

Search completed: July 15, 2005, 15:41:45
Job time : 1226 secs

Result No.	Score	Query		Length	DB	ID	Description
		Match					
C 1	273	14.1	300	1	AV178298	AV178298	AV178298
C 2	124.6	6.4	476	4	B1741875	B1741875	AV178298
C 3	57.4	3.0	1101	9	CNS0039G	CNS0039G	AV178298
C 4	55.8	2.9	478	6	CNS005TE	CNS005TE	AV178298
C 5	54.4	2.8	476	6	CD167558	CD167558	AV178298
C 6	54.4	2.8	534	6	CD185097	CD185097	AV178298
C 7	53.8	2.8	1359	9	AG310736	AG310736	AV178298
C 8	53.6	2.8	622	5	BU418103	BU418103	AV178298
C 9	53	2.7	488	8	AZ379487	AZ379487	AV178298
C 10	53	2.7	1196	9	CNS0207H	CNS0207H	AV178298
C 11	50.8	2.6	610	9	CR089470	CR089470	AV178298
C 12	50.8	2.6	838	9	CR214149	CR214149	AV178298
C 13	50.8	2.6	1101	9	CNS0037Q	CNS0037Q	AV178298
C 14	50.8	2.6	1133	9	CL111125	CL111125	AV178298
C 15	50.6	2.6	1201	9	CNS0203O	CNS0203O	AV178298
C 16	50.2	2.6	440	8	AZ408774	AZ408774	AV178298
C 17	50.2	2.6	572	8	AZ959862	AZ959862	AV178298
C 18	50	2.6	787	9	AG585416	AG585416	AV178298
C 19	49.8	2.6	550	8	BH264512	BH264512	AV178298
C 20	49.6	2.6	501	6	CD167504	CD167504	AV178298
C 21	49.6	2.6	562	9	CB832047	CB832047	AV178298
C 22	49.2	2.5	637	9	CE379668	CE379668	AV178298
C 23	49	2.5	744	9	AG447497	AG447497	AV178298
C 24	48.8	2.5	835	8	BZ206528	BZ206528	AV178298

QY 1718 GAAGTCGCTCATCTCCCATCTCTTGATCCAACTATGCAGCGCTTGAAACACGCGTTCT 1777
 Db |||||
 240 GAAGTCGCTCATCTCCCATNTCTTGATCCAATATGCCANCTTGAAACACGCGTTCT 181
 QY 1778 CGACTCAGGGCGAGCCGCTGACGAGTGGGATACCATGTTCCAAATCCCAATAACTGGA 1837
 Db |||||
 180 CGACTCAGGGCGAGCCGCTGACGAGTGGGATACCATGTTCCAAATCCCAATAACTGGA 121
 QY 1838 ATCTCGAGATGAAGTTGGGTTCAAAATGAACCTCAAAAACAAACGATTCGTTGGAGGAA 1897
 Db |||||
 120 ATCTCGAGATGAAGTTGGGTTCAAAATGAACCTCAAAAACAAACGATTCGTTGGAGGAA 61
 QY 1898 ATGGAGCATTTGATATGCTGCTGCACTGGGATGTAG 1932
 Db |||||
 60 ATGGAGCATTTGATATGCTGCTGCACTGGGATGTAG 26

RESULT 2
 BI741875 476 bp mRNA linear EST 25-SEP-2001
 LOCUS Strongyloides ratti L1 pAMP1 v3 Chiapelli McCarter
 DEFINITION Strongyloides ratti cDNA 5' similar to TR:P91120 P91120 SIMILAR TO HMG-BOX SINGLE-STRANDED DNA AND RNA BINDING PROTEINS. [1] ;, mRNA sequence.
 ACCESSION BI741875
 VERSION BI741875.1 GI:15742831
 KEYWORDS EST.
 SOURCE Strongyloides ratti
 ORGANISM Strongyloides ratti
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Panagrolaimoidea; Strongyloidea; Strongyloidea.

REFERENCE 1 (bases 1 to 476)
 AUTHORS McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J.,
 Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B.,
 Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C.,
 Tsagareishvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C.,
 Underwood,K., Stepcoe,M., Allen,M., Person,B., Swaller,T.,
 Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
 McCann,R., Waterston,R. and Wilson,R.
 TITLE The Washington Univ. Nematode EST Project, 1999
 JOURNAL Unpublished (1999)
 COMMENT Contact: McCarter JP
 The Washington Univ. Nematode EST Project, 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

The library was constructed by Brandi Chiapelli and Dr. James McCarter (bchiapell@watson.wustl.edu & jmcarter@watson.wustl.edu) at Washington University, St. Louis. DNA Sequencing by: Washington University Genome Sequencing Center St. Louis.
 Seg primer: -40RP from Gibco
 High quality sequence stop: 306.
 Location/Qualifiers

FEATURES
 source
 1..476
 /organism="Strongyloides ratti"
 /mol_type="mRNA"
 /db_xref="taxon:34506"
 /dev_stage="L1"
 /lab_hosts="DH10B"
 /clone_lib="Strongyloides ratti L1 pAMP1 v3 Chiapelli McCarter"
 /note="vector: pAMP1 (Gibco); The library was constructed by Brandi Chiapelli and Dr. James McCarter at Washington University, St. Louis. The cDNA was made by using Dynabead oligo-dT priming (Dynal). PCR based library using a modified protocol from the SMART PCR cDNA Synthesis Kit from Clontech. Directionally cloned into the UDG sites of pAMP1. Nematodes were provided by Dr. Mark Viney of Bristol, UK."

ORIGIN

Query Match 6.4%; Score 124.6; DB 4; Length 476;
 Best Local Similarity 63.5%; Pred. No. 1.3e-21;
 Matches 209; Conservative 0; Mismatches 114; Indels 6; Gaps 1;
 QY 1523 GTAGAACACGCTGTCAAAAACAAATTCGTGCATTTACCAAGACTGTTTCGGTGC-----ACCAA 1576
 Db |||||
 85 GTGTACACGCTGGAATTCGAATTTACACATATCCAGTATGTATGCTGCTAAATTTATTTG 144
 QY 1577 CTCACAAAGCTGAAATGATGTATGCAAGAGTATTTCCAAAGATATTGAAGAACGTCCTCATTC 1636
 Db |||||
 145 GTGATGAACCAATATTTATGATAATAAACAACAAATTTGTTGAGGAAAGAGAAATTC 204
 QY 1637 CTCATTGTTCTTTGAACCAAAAGGAGCATACGATTCAGATTCGACTGGAGCAATG 1696
 Db |||||
 205 CTCACATTTTTTTTGTACCAAAAGGAAACATACAAAGATTAAAGTGGGTAAACAGCAACAG 264
 QY 1697 AAAAGAAATTCAGGACTTTGGAAGTCGTTTCATTTCTCCCATCTCTTTGATCCCAACTATGC 1756
 Db |||||
 265 AACAGAAATTCGCGAATTTGGAAGTAGATTATATACCATCATCTTATCCAACTAGGC 324
 QY 1757 CAGCTTGAACACGGCTTTCTGCACTCAGGGGCGAGCCCTGACGAGTGGGATACCATGT 1816
 Db |||||
 325 CAGCAATTAATTCAGTTGTATCAACACAGGAAAGAAAGAAATGAATATATGAACCAACTT 384
 QY 1817 TCAAAATCCGAAATAACTGGAATCCCTGGA 1845
 Db |||||
 385 GGAATAATACCAATGCTTGGAAATCTGGA 413

RESULT 3
 CNS00396/c
 LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Direct Submission

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Osoegawa and

Aaron Mammose in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

NY. The library is named RPCI-98 and was constructed by partial

Scori digestion of Drosophila DNA provided by the BDGP from the

isogenic strain y2; cn bw sp, the same strain used for the BDGP's

P1 and EST libraries. A more detailed description of the library

and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be

found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

1..1101

/organism="Drosophila melanogaster"

/mol_type="genomic DNA"

/db_xref="taxon:7227"

/clone_lib="BACRPCI-98"

/note="end : TET3"

ORIGIN

Query Match	3.0%	Score 57.4	DB 9	Length 1101
Best Local Similarity	15.3%	Pred. No. 0.0011		
Matches 92	Conservative 262	Mismatches 247	Indels 0	Gaps 0
Qy	677	GAGGTGAAGAACCAACGCGATGATGAATCGATGGGACCAAGAAATCCAAAGAGAGAAC	736	
Db	1100	KAREWGDDTWDRDTRFKDDWDKWKWTTWKORADDRRWAGDADRWWDGAGTWTWTATWW	1041	
Qy	737	TTCAGAGGAGAGATGAAGATGATGATCTTGAGGACGAGGATGTACCCAGAGAAGAA	796	
Db	1040	WWWWTATWTWDKWWWWATAAKTDTATWMTATWRADWAGRDGAKGRDRAADADGA	981	
Qy	797	GTTCGGATGGAGAACCAAGTCAACGAGCATCAGAGAGAGATTTAGCCAGGAGAT	856	
Db	980	GRRDGGRKKRDKDKRDKDGDGDKKGGKKKKAKAAKWATKWDWDWDKWKWDGAKDRKA	921	
Qy	857	TGAAAAGTAGTCTAGAGTATAAAGAGCTTTTACAGAAATGCGGAAGTTCAATCATCTGCT	916	
Db	920	DDDDGAGDKDDGKGKDADDDTGTDKDDDKDWDWDKAKGTWGDATWAAATDWWWG	861	
Qy	917	CTTACCAACGAATGAGGAGATTCTCCACTGAGCAAGCAAGCCCTTTGGCTATGAACGATG	976	
Db	860	WADADWTTWDAADDDWADDWDADAWAKWDDADAWAGARTADRRDWDGDRKGRGAKRRD	801	
Qy	977	AGGATGAAGTGCAATTCCGCGCAATGAGGCTCGTGCAAAACTAGATCAAAAATCTCAAC	1036	
Db	800	RKRADDKRDADDRDDAATWTTWTTTRDTDMDWKWKTDTWTRWADRTWDRDDDDDRDR	741	
Qy	1037	TTTGGCTCGGTCTCCATGGTTTTGGAGAGTCTGATGATGATGAAGCAAGAAGATGAAA	1096	
Db	740	AGTAGRKWRRTWKBRWKRBDTRWDDADADTARDRRRGDGDAGKAGKKTGRKRRRD	681	
Qy	1097	ATTTGATGATCCATCTGAAAATTCATCCGTGTCGACCACTTCGCTTTCCTCCGGAT	1156	
Db	680	RATWDRTDAAWADAWWWTTTDTDDTDWDXDRRRKGRARRRRTTARAAMDWMTKAWDWA	621	
Qy	1157	TCGTTGAGAATTAAGATCAAAATGATGAAATTTGGAAGTGCATTGGACAGAAATTAATATC	1216	
Db	620	KWDWKTRADRWDRWAADTWDTDARKADRDWAKARAWRARRDRARAARDRWTTKGKTFTA	561	
Qy	1217	GAGTTGATGACGTGGAAAAGTATCTTCTCCAAAGCCGATGGAATTCAAATCCAAAACCTC	1276	
Db	560	TWTTWAAAANWAAWAAATTAWTTTWTWTTTWTWTTTWTWTTTAAWAAWWTATW	501	
Qy	1277	A	1277	
Db	500	A	500	

RESULT 4
CNS005TE/c
LOCUS
DEFINITION
997 bp DNA linear GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR12K22 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
VERSION AL060767.1 GI:4943573
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 997)
REFERENCE
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazuhiro Oosogawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial, EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <http://bacpac.med.buffalo.edu/drosophila/bac.htm>.

FEATURES

```

location/qualifiers
1. .997
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR12K22"
/clone_lib="RPCI-98"
/note="end : TET3"

```

ORIGIN

Query Match 2.9%; Score 55.8; DB 9; Length 997;
Best Local Similarity 18.6%; Pred. No. 0.0029;
Matches 79; Conservative 154; Mismatches 192; Indels 0; Gaps 0;

[illegible]

RESULT 5	
CD167558/c	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	

CD167558 478 bp mRNA linear EST 14-SEP-2003
MM1-0013G-V027-H10-U.B MM1-0013 Schistosoma mansoni cDNA clone
MM1-0013G-V027-H10.B, mRNA sequence.

ACCESSION CD167558
 VERSION CD167558.1
 KEYWORDS EST.
 SOURCE Schistosoma mansoni
 ORGANISM Schistosoma mansoni

ORGANISM. *Schistosoma mansoni*
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigeida; Schistosomatoidea; Schistosomatidae; Schistosoma.
1 (bases 1 to 478)
REFERENCE
AUTHORS Vertovskii, A. M.; Almeida, S.; DeMarco, R.; Martins, E. A. L.; Guimaraes, P. E. M.

TITLE	COMMENT
<p>Direct Submission Submitted (12-APR-2000) Genoscope - Centre National de Sequençage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr) - Web : www.genoscope.cns.fr</p>	<p>This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetradon.</p>

```

FEATURES
source
    map//.../genome/pep/22/Tetraodon...
    Location/Qualifiers
        1..1196
            /organism="Tetraodon nigroviridis"
            /mol_type="genomic DNA"
            /db_xref="taxon:99883"
            /clone="I52P14"
            /clone_lib="G"
            /note="Genoscope sequence ID : COAG152DH07LP1-end : T7"

```

```

ORIGIN
Query Match      2.7%; Score 53; DB 9; Length 1196;
Best Local Similarity 52.0%; Pred. No. 0.017;
Matches 119; Conservative 0; Mismatches 110; Indels 0; Gaps 0;
/note= genomecope sequence ID : CGAG132D00/DF1-end : 17/

```

653	Qy	AAATGCCCCAGTTGCCGTGGAAGAGTGTAAGAAACAACGAGATGATGATGAATCGAGTGG	712
120	Db	AGAACCCGCGAGCGGGAAGAAGAAGAAGAAAGAAAGAGAGAAGAAGAAGAAGA	179
713	Qy	ACCAAGAATGCAACAAACAGAACTTCAAGACGGAAGATGAAGATGATGATCTTTGAGG	772
180	Db	AGAGAAGAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGA	239
773	Qy	ACGAGGATGTACCCAGAAAGAAAGTTTCGGATTGGAGACCACAAAGTGAAGCAGAGCATC	832
240	Db	AGNAGAAAGAGAGAAAGAAAGNAGNAGNAGNAGNAGNAGNAGNAGNAGNAGNAGAGA	299
833	Qy	AGAGAAGAGATTTAGCCAGGAGATTGAAAAGTAGTCTTAGATTTAAAAAGA	881
300	Db	AGAGAAGAGAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA	348

RESULT 11	CR089470	610 bp	DNA	linear	GSS 05-JUL-2004
LOCUS	CR089470/c	Reverse strand read from insert in 5'HPT insertion targeting and			
DEFINITION	chromosome engineering clone MHPN28ii7, genomic survey sequence.				
ACCESSION	CR089470				
VERSION	CR089470.1	GI:49823062			
KEYWORDS	GSS; genome survey sequence; MICER.				

SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. 1 (bases 1 to 610)
AUTHORS	Adams,D.J., Biggs,P.J., Cox,A.V., Davies,R.M., van der Weyden,L., Jonkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,I., Yu,Y., Rogers,J. and Bradley,A.

TITLE	Direct Submission
JOURNAL	Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. http://www.sanger.ac.uk/MICR
FEATURES	Location/Qualifiers
Source	1..610 /organism="Mus musculus" /mol_type="genomic DNA" /db_xref="taxon:10090" /clone="MHF28117" /db_xref="taxon:10090"

ORIGIN	/clone_11b="MHFN"			
	Query Match	2.6%;	Score 50.8;	DB 9; Length 610;
	Best Local Similarity	55.7%;	Pred. No. 0.054;	
	Matches 97; Conservative	0;	Mismatches 77;	Indels 0; Gaps 0;
Qy	682	GAAGAACACCGGATGATGATGATCGATGGACCAAGAANTGCACAAAGAGAACTTCAA	741	

Qy 682 GAAGAACACGGATGATGATGATCGAGTCGACCAAGAATGCAACAAGAGACTTCAA 741

497	GAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA	438
Db		
742	GAGGAAGATGAAGATCATGATCTTTGAGGACGAGGATGTACCCAGAAAGAAAGTTCG	801
QY		
437	GAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA	378
Db		
802	GATGGGAAGACCAAAAGTGAAGCAGAGCATCAGAGAAAGAGATTTAGCCAGGAGA	855
QY		
377	AAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG	324
Db		

RESULT 12	
CR214149/c	
LOCUS	CR214149
DEFINITION	Reverse strand read from insert in 5'HPRT insertion targeting and chromosome engineering clone MHPN328h07, genomic survey sequence.
ACCESSION	CR214149
VERSION	CR214149.1
KEYWORDS	GSS; genome survey sequence; MTCER.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
	linear DNA 838 bp
	GSS 06-JUL-2004

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 838)
 REFERENCE
 AUTHORS
 Adams, D.J., Biggs, P.J., Cox, A.V., Davies, R.M., van der Weyden, L.,
 Jonkers, J., Smith, J., Plumb, R.W., Taylor, R.G., Nishijima, I., Yu, Y.,
 Rogers, J. and Bradley, A.
 Direct Submission
 TITLE
 JOURNAL
 Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. <http://www.sanger.ac.uk/MICER>
 FEATURES
 Location/Qualifiers
 1. .838

```

/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="MHPN328h07"
/clone_lib="MHPN"
ORIGIN

```

	Query Match	2.6%	Score 50.8	DB 9	Length 838
	Best Local Similarity	55.7%	Pred. No. 0.059		
	Matches 97	Conservative 0	Mismatches 77	Indels 0	Gaps 0
QY	682	GAAGAACACGATGATGATCGAGTGGACCAAGAATGCCAACAAAGAGACTTCAA	741		
Db	497	GAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA	438		
QY	742	GAGGAAGATCGAAGATGATCATCTTTGAGGACGAGGATGTACCCAGAAGAGAACTTCG	801		
Db	437	GAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA	378		
QY	802	GATGAGAACCCAAAGTGAACGACGACATCAGAGAAGAGATTTAGCCAGGAGA	855		
Db	377	AAAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA	324		

RESULT 13	CNS0037Q	linear	GSS 03-JUN-1999
LOCUS	CNS0037Q	1101 bp	
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K14 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.		
ACCESSION	AL064465		
VERSION	AL064465.1	GI:4941722	
KEYWORDS	GSS.		
SOURCE	Drosophila melanogaster (fruit fly)		
ORGANISM	Drosophila melanogaster		
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
REFERENCE	1 (bases 1 to 1101)		
AUTHORS	Genoscope.		
TITLE	Direct Submission		


```
OY 737 TTCAAGAGGAGATGAAGATGATGATGATCTTTGAGGACGAGGATGTACCCAGAGAAGAA 796
Db 593 ANNNNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAAN 534
OY 797 GTTCGATGGAGACCAACAAAGTGAAGCAGAGCATCAGAGAAGAGATTGACCGAGAGAT 856
Db 533 NNANNNNNAANNAANNAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 474
OY 857 TGAAGAAGTAGCTAGATTAAAGAGCTTTTACAGAAATGCGGAAGTTCAATCATTTGCTCT 916
Db 473 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 414
OY 917 CTTACCAACGATGAGGAGATTCTCCACTGACCAAGCGAAGCCCTTTG 963
Db 413 NNNNAANNAANNAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 367
```

RESULT 15
CNS02030/c
LOCUS
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
152B16 of library G from Tetraodon nigroviridis, genomic survey
sequence.

ACCESSION AL206205
VERSION AL206205.1 GI:7865024
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis
ORGANISM Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorphi; Acanthopterygii; Percomorphi; Tetraodontiformes;
Tetraodontoidea; Tetraodontidae; Tetraodon.

REFERENCE

1 Roest Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,
Saurin, W. and Weissenbach, J.
Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
Nat. Genet. 25 (2), 235-238 (2000)

REFERENCE

2 Roest Crolius, H., Jaillon, O., Dasilva, C., Ozouf-Costaz, C.,
Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F.,
Saurin, W., Bernot, A. and Weissenbach, J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)

JOURNAL

MEDLINE
PUBMED

20359837
10899143

3 (bases 1 to 1201)

REFERENCE

Genoscope.

Direct Submission

Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : secref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)

This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
<http://www.genoscope.cns.fr/tetraodon>.

FEATURES

source

1. 1201
/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99863"
/clone="152B16"
/clone_lib="G"
/note="Genoscope sequence ID : COAG152DA08LP1-end : T7"

ORIGIN

Query Match

Best Local Similarity 2.6%; Score 50.6; DB 9; Length 1201;

Pred. No. 0.074;

Matches 110; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

```
OY 673 GGAAGAGGTGAAGAAACAACGCGATGATGATGATCGAGTGGACCAAGAAATCCAAACAAGA 732
Db 1194 GAAGAAGAGNAGNAGNAGNAGNAGNAGNAGNAGNAGNAGNAGNAGNAGNAGNAGNAG 1135
OY 733 GAACCTTCAAGAGGAGAGATGAAGATGATGATCTTTGAGGACGAGGATGTACCCAGAAGA 792
Db 1134 AGAAGAGAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1075
OY 793 AGAAGTTCCGATGCGAGAACCAACAAAGTGAACGACAGCATCAGAGAAGAGATTTAGCCAGG 852
Db 1074 AGAAGNAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1015
OY 853 AGATTGAAAAGTAGTCTCTAGATTAAAAAGA 881
Db 1014 AGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 986
```

Search completed: July 15, 2005, 14:57:35
Job time : 6512 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 18, 2005, 12:25:30 ; Search time 7749 Seconds
(without alignments)
4020.739 Million cell updates/sec

Title: US-10-736-868-2
Perfect score: 3251
Sequence: 1 MILEFLFLLLGFCIAPLSA.....SKTRFVGCGGAFDMPALGL 643

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2.1/USPTO.spool/US10736868/runat.14072005.105509.2278/app.query.fasta.1.839
-DB=GenEmbl -QWMT=fastap -SUFFIX=rgc -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORES=ptc -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10736868 @CNC 1.1.4545 @runat.14072005.105509.2278 -NCPU=6 -ICPU=3
-NO.MWAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	3291	100.0	2131	3	AY360470 Caenorhab
C 2	2938	89.3	45389	3	U80032 Caenorhabdi
C 3	1873.5	56.9	17418	3	AC084485 Caenorhab
C 4	585	17.8	39089	3	CBRG33D04

5	181	5.5	3631	6	CQ575340	Sequence
6	181	5.5	3745	3	AF367177	Drosophila
7	180	5.5	5894	3	ACMHC	Acanthamoeba
C 8	179.5	5.5	19514	8	SPAC29E6	S. pombe chr
C 9	179.5	5.5	36315	8	SPAC30	S. pombe c
10	177	5.4	3543	6	AX151655	Sequence
C 11	177	5.4	292967	14	AF369029	White spo
C 12	177	5.4	305107	6	AX151396	Sequence
C 13	177	5.4	305107	14	AF332093	White spo
14	176.5	5.4	7396	6	AX686226	Sequence
15	175.5	5.3	5122	9	HUMYONM	Human nonmu
16	174	5.3	6788	5	XELNMUSMYO	Xenopus lae
17	173.5	5.3	6974	5	AF055895	Xenopus l
18	173	5.3	7450	9	AB191263	Homo sapi
C 19	173	5.3	307287	14	AF405070	Homo sapi
20	172	5.2	5670	3	TCA306290	Shrimp wh
C 21	171.5	5.2	110000	8	AE016815_4	Toxocara
22	171	5.2	96210	1	YPPMT1	Continuation (5 of
C 23	171	5.2	100984	1	AF053947	Yersinia
24	171	5.2	100990	1	AF074611	Yersinia
25	171	5.2	100990	6	AR487637	Yersinia
C 26	171	5.2	106642	1	AE017045	Sequence
27	170.5	5.2	4281	9	HSMYOIHMA	AE017045
28	170.5	5.2	5556	9	AK131080	H. sapiens m
29	169.5	5.2	2896	10	BC044834	Homo sapi
30	169.5	5.2	5254	10	AK131171	Mus muscu
31	169.5	5.2	5946	6	CQ730200	Mus muscu
32	169.5	5.2	6137	9	AE002376	Human msn
33	169.5	5.2	68727	3	AC004516	Sequence
34	169	5.1	6499	3	AF136711	Drosophila
35	168.5	5.1	3858	6	CQ735115	Amoeba pr
36	168.5	5.1	7355	10	MMU312390	Sequence
37	168.5	5.1	53424	3	AB055861	Mus muscu
C 38	168.5	5.1	184704	2	CB385063	Procamb
39	168	5.1	1868	5	BC083371	Danio rer
40	168	5.1	5439	6	CQ573099	Sequence
41	168	5.1	5756	6	CQ575339	Sequence
42	168	5.1	30386	2	AC014301	Sequence
C 43	168	5.1	134053	8	BX908812	Neurospor
C 44	168	5.1	185405	3	AC023743	Drosophila
C 45	168	5.1	299686	3	AE003444	Drosophila

ALIGNMENTS

RESULT 1	AY360470	2131 bp	linear	INV 01-JUN-2004
AY360470	Caenorhabditis elegans osmotic stress resistance protein (osr-1)			
LOCUS	mRNA, osr-1-rm1 allele, complete cds.			
DEFINITION	AY360470			
ACCESSION	AY360470.1	GI:38154581		
VERSION				
KEYWORDS	Caenorhabditis elegans			
SOURCE	Caenorhabditis elegans			
ORGANISM	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;			
	Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.			
REFERENCE	1 (bases 1 to 2131)			
AUTHORS	Solomon, A., Bandhakavi, S., Jabbar, S., Shah, R., Beitel, G.J. and Morimoto, R.I.			
TITLE	Caenorhabditis elegans OSR-1 Regulates Behavioral and Physiological Responses to Hyperosmotic Environments			
JOURNAL	Genetics 167 (1), 161-170 (2004)			
PIUMED	15166144			
REFERENCE	2 (bases 1 to 2131)			
AUTHORS	Solomon, A., Bandhakavi, S., Jabbar, S., Shah, R., Beitel, G.J. and Morimoto, R.I.			
TITLE	Direct Submission			
JOURNAL	Submitted (04-AUG-2003) Biochemistry, Northwestern University, 2205 Tech Drive, Hogan 2-100, Evanston, IL 60201, USA			
FEATURES	Location/Qualifiers			
source	1. .2131			
	/organism="Caenorhabditis elegans"			

```

/mol_type="mRNA"
/db_xref="taxon:6239"
/chromosome="I"
/map="0.2 m.u."
1. .2131
/gene="osr-1"
/allele="osr-1-rml"
1. .98
/gene="osr-1"
/allele="osr-1-rml"
99. .2030
/gene="osr-1"
/allele="osr-1-rml"
/function="mediates multiple responses to osmotic stress"
/notes="OSR-1"
/codon_start=1
/product="osmotic stress resistance protein"
/protein_id="AAB1229.1"
/db_xref="GI:38154582"
/translations="MILFLFLLLGFCIAPLSAQSPSTSDAPGALLSLVKGSHOKL
PLAPSMELALMGVQFVDALIKKGOMEMAKGAFKQLLEVLKVPDQFDYKKLKYDD
LAADAVYQAEAKLQPKSGNAFIDMLNGIPIGSIRGLEDAITQDMENTDPSE
OIAKAVMDKFOIILPGLVANMIAGNPKPMQOKAKAAPSSVFQALAOORAMLGK
NAPVAGRGEEORSMNVRDORMQORELOEDEDDEDDLEDEVPRESSSDGEPSAE
HQRRDLARLKSSPRLEKLLONAEVQSLLSYQMRDSPLSKRPPLANNDEDESAFRM
EAKALDQKSLVLGLHGFGEDEDDDEENLIDPSNSFRAPRLRLSSGFVEKLKS
NDELKSLDRIKYRVDVDEKYLAPKPMENPKPGYFAPRKIPTRPKWMLPILIGSD
PKVOEIRHPSTEWIKESRVLTNLKNPISLAALFMDDKLENTLKRQMLTDEQKG
RTRVKTRIALPRFLGAPTAKEMIDAKVFODIEERPIPLFPFPKGRHRLRWTGANE
KEIPGLGSRFLPSLDPTPALNTAFSTQGRARDEWDTFPKIPNNWNPGEVGFKNKS
KTRRFVGGNGAFDMPALGL"
2031. .2131
/gene="osr-1"
/allele="osr-1-rml"

ORIGIN
Alignment Scores:
Pred. No.: 1,33e-205 Length: 2131
Score: 3291.00 Matches: 643
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-736-868-2 (1-643) x AY360470 (1-2131)

QY 1 MetIleLeuPheLeuPheLeuLeuGlyPheCysIleAlaProLeuSerAla 20
DB 99 ATGATTTATTTTATTTTATTTTATTTTCTCTGTTAGGATTTTGTATCGCACCATTATCGGCC 158
QY 21 GlnSerProSerThrSerAspAlaProGlyAlaLeuLeuSerSerLeuValGlyLysSer 40
DB 159 CAATCTCATCGACTCCGATGCTCCGGAGCTTGTGTGTCATCTCTCGTAGGTAAAGC 218
QY 41 HisGlnLysLeuProLeuAlaProSerMetGluAlaLeuLeuMetGlyValGlnPhe 60
DB 219 CATCAAAATACCACTGGCTCCATCAATGAAAGCTTGAAGTGTGGTGTTCATTT 278
QY 61 ValAspAlaLeuLeuLysGlyGlnMetGluMetAlaLysGlyAlaPheLysThrGln 80
DB 279 GTTGATGCTCTCATCAAAAAGGTCAATGGAAATGGCAAAAGGAGCATTTAAGACTCAA 338
QY 81 LeuGluValLeuGluLysValHisProAspGlnPheAspLysTyrLysLysLeuLysVal 100
DB 339 TTAGAAGTCTTAGAGAAAGTACATCTCTGATCAATTCGATAAGTACAAAAGCTAAAAAGTT 398
QY 101 AspAspLeuAlaAlaAspAlaValMetGlnAlaGluMetAlaLysLeuGlnProLys 120
DB 399 GATGATTGGCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 458
QY 121 SerGlyAsnAlaPheIleAspMetLeuAsnGlyIleProIleGlySerSerIle 140
DB 459 TCAGGAATGCATTTATTCGATATGTTGATGATGATGATGATGATGATGATGATGATGAT 518

```

QY 501 MetLeuThrAspGluGlnIysGlyArgThrArgValIysThrIleArgAlaLeuProArg 520
 |||||
 Db 1599 ATGTTAACTGATGAACAGAAAGGTAGAACCGTGTCAAAACAATTCGTGCATTACCAAGA 1658
 |||||
 QY 521 LeuPheGlyAlaProThrAlaLysAlaGluMetIleAspAlaLysValPheGlnAspIle 540
 |||||
 Db 1659 CTGTTGGTGCACCAACTGCAAAAGCTGAATGATGTGCAAGGTATTCAGATATT 1718
 |||||
 QY 541 GluGluArgProIleProPhePheGluProLysGlyArgIleAspGlyArgLeuArg 560
 |||||
 Db 1719 GAAGAACGTCCTCCATTCCTGCTTTTGAACCAAAAGGAGCATACAGATTGAGA 1778
 |||||
 QY 561 TrpThrGlyAlaAsnGluIysGluIleProGlyLeuGlySerArgPheIleLeuProSer 580
 |||||
 Db 1779 TGGACTGGAGCAAAATGAAGAAATTCAGGACTTGGAGCTTCGATTCCTCCATCT 1838
 |||||
 QY 581 LeuAspProThrMetProAlaLeuAsnThrAlaPheSerThrGlnGlyArgAlaArgAsp 600
 |||||
 Db 1839 CTTGATCCACTATGCCAGCCTTGAAACACGGCTTCTCGACTCAGGGGCGAGCCGTCAC 1898
 |||||
 QY 601 GluTrpAspThrMetPheLysIleProAsnAsnTrpAsnProGlyAspGluValGlyPhe 620
 |||||
 Db 1899 GAGTGGGATACCATGTTCAATCCCGAATCACTGATCTCTGGAGATGAAGTTGGGTTC 1958
 |||||
 QY 621 LysMetAsnSerLysThrLysArgPheValGlyGlyAsnGlyAlaPheAspMetProAla 640
 |||||
 Db 1959 AAAATGAATCAAAAACCAACGATTCGTGGAGGAATGGAGCATTTGATATGCTGCA 2018
 |||||
 QY 641 LeuGlyLeu 643
 |||||
 Db 2019 CTGGGATTG 2027
 |||||
 RESULT 2
 U80032/c
 LOCUS U80032 45389 bp DNA linear INV 16-SEP-2004
 DEFINITION Caenorhabditis elegans cosmid C32E12, complete sequence.
 ACCESSION U80032
 VERSION U80032.1 GI:2073526
 KEYWORDS HTG.
 SOURCE Caenorhabditis elegans
 ORGANISM Caenorhabditis elegans
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditodea; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 45389)
 WormBase Consortium
 Genome sequence of the nematode C. elegans: a platform for
 investigating biology. The C. elegans Sequencing Consortium
 Science 282 (5396), 2012-2018 (1998)
 JOURNAL 99069613
 MEDLINE 9851916
 PUBLISHED 9851916
 REFERENCE 2 (bases 1 to 45389)
 Wilcox,L.
 The sequence of C. elegans cosmid C32E12
 JOURNAL Unpublished (2001)
 REFERENCE 3 (bases 1 to 45389)
 Waterston,R.
 Direct Submission
 JOURNAL Submitted (27-NOV-1996) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 REFERENCE 4 (bases 1 to 45389)
 Waterston,R.
 Direct Submission
 JOURNAL Submitted (08-MAY-1997) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 REFERENCE 5 (bases 1 to 45389)
 Waterston,R.
 Direct Submission
 JOURNAL Submitted (18-OCT-2001) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

6 (bases 1 to 45389)
 Waterston,R.
 Direct Submission
 Submitted (19-APR-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 7 (bases 1 to 45389)
 Waterston,R.
 Direct Submission

REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (20-JUL-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 8 (bases 1 to 45389)
 Waterston,R.
 Direct Submission

REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (19-NOV-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 9 (bases 1 to 45389)
 Waterston,R.
 Direct Submission

REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (13-JAN-2003) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 10 (bases 1 to 45389)
 WormBase Consortium
 Direct Submission

REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (16-SEP-2004) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 On May 8, 1997 this sequence version replaced gi:1703546.
 Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1RQ, England
 email: submissions@watson.wustl.edu and jes@sanger.ac.uk

COMMENT

NOTICE: This sequence may not be the entire insert of this clone.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we provide a small overlap between
 neighboring submissions.

This sequence was finished as follows unless otherwise noted: all
 regions were double stranded, sequenced with an alternate chemistry
 or covered by high quality data (i.e., phred quality >= 30); an
 attempt was made to resolve all sequencing problems, such as
 compressions and repeats; all regions were covered by sequence from
 more than one m13 subclone.

For a graphical representation of this clone sequence and its
 analysis see:
<http://www.wormbase.org/db/seq/sequence?name=C32E12;class=Sequence>

NEIGHBORING CLONE INFORMATION

The 5' clone is F56A3, 200 bp overlap; the 3' clone is B0261, 3201
 bp overlap. Actual start of this clone is at base position 1 of
 C32E12; actual end is at 42388 of C32E12.

NOTES:

Coding sequences below are the result of integration and manual
 review of the following data : computer analysis using the program
 Genefinder (P. Green and L. Hillier, personal communication), the
 large scale EST projects of Yuji Kohara
 (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C.
 elegans ORFome cloning project (<http://worfdb.dfci.harvard.edu/>),
 similarity to other proteins from Blastx analyses
 (<http://blast.wustl.edu/>), sequence conservation with C. briggsae
 using Jim Kent's WABA alignment program (Genome Research

Db 3416 CCAGCCCTTGACCGGCTTCTCGACTCAGGGGAGCCCGTGACGAGTGGGATACCATG 3357
Qy 606 PhelysileProAsnAsnTrpAsnProGlyAspGluValGlyPhelysMetAsnSerLys 625
Db 3356 TTCAAAATCCGAATAACTGGAATCCTGGAGATGGAAGTTGGTTCAAAATGAACTCAAAA 3297
Qy 626 ThrlysrArgPheValGlyGlyAsnGlyAlaPheAspMetProAlaLeuGlyLeu 643
Db 3296 ACCAAACGATTCGTTGGAGAAATGGAGCAATTTGATATGCTGCACCTGGGATTC 3243

RESULT 3

CBRG03012/c 17418 bp DNA linear INV 04-NOV-2000
LOCUS CBRG03012
DEFINITION Caenorhabditis briggsae cosmid G03012, complete sequence.
ACCESSION AC084485
VERSION AC084485.1 GI:11094935
KEYWORDS HTG.
SOURCE Caenorhabditis briggsae
ORGANISM Caenorhabditis briggsae
REFERENCE 1 (bases 1 to 17418)
AUTHORS Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
TITLE The C. briggsae Genome Sequencing Project.
JOURNAL Washington University Genome Sequencing Center.
REFERENCE 2 (bases 1 to 17418)
AUTHORS Unpublished
TITLE The sequence of C. briggsae cosmid G03012
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 17418)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (04-NOV-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA
e-mail: jpeith@watson.wustl.edu

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

FEATURES

source
1. 17418
/organism="Caenorhabditis briggsae"
/mol_type="genomic DNA"
/strain="Gujarat G16"
/db_xref="taxon:6238"
/clone="G03012"

ORIGIN

Alignment Scores:
Pred. No.: 7e-112 Length: 17418
Score: 1873.50 Matches: 522
Percent Similarity: 28.06% Conservative: 31
Best Local Similarity: 26.48% Mismatches: 24
Query Match: 56.93% Indels: 1396
DB: 3 Gaps: 11

US-10-736-868-2 (1-643) x CBRG03012 (1-17418)

Qy 62 AspAlaLeuIleLysLysGlyGlnMetGluMetAlaLysGlyAlaPheLysThrGlnLeu 81
Db 17417 GATGCTTTAATTAATAAAGACAGATGGAATGGCAAAAGAGCATTCAAAAACACACTG 17358
Qy 82 GluValLeuGlyLysValHisProAspGlnPheAspLysTyrLysLysLeuLysValAsp 101
Db 17357 GAGGTTCTGGAGAAAGTACATCCGGATCAATTTTGAGAAATATAAGAACTGAAAGTGGAG 17298

Qy 102 AspLeuAlaAlaAspAlaValMetGlnGlnAlaGluMetAlaLysLeuGlnProLys--- 120
Db 17297 GATTTAGCGCAGATCAGTTATGCAACAGGCAGGAATGTCCAAGTTGCAACCGCAAACT 17238
Qy 120 ----- 120
Db 17237 GGTAAAGGAGAAATATATAAAGAAGATCATTTACTACGAACAATACGACTGAAAATTAATT 17178
Qy 120 ----- 120
Db 17177 TCTAGTTAGCACATGCAACAGTAAAAAATTTCAAAATTTCAAGTTTTCAGTTTTCAGATAATGTCAGAA 17118
Qy 120 ----- 120
Db 17117 AAAGTTTTTATTTTCATTTTAAATTAATGTAAGTATCAATTTCCGATCGAATCAATTTG 17058
Qy 120 ----- 120
Db 17057 GCAATCATTTAGAGGAATAAGTTTCTCGAAACTGAAAGTTTTCAAATGTCAGATCTTG 16998
Qy 120 ----- 120
Db 16997 AAACGAACTTTGAAAGTCTGAGAAACCTTAAAGTTAAACATTTGAACACTGTCTAATAG 16938
Qy 120 ----- 120
Db 16937 ACAGGCGTCTTAACCTTCAAGTTTTCGAGACTTTTATATTTTGGTGAAAAATGAGTAATA 16878
Qy 120 ----- 120
Db 16877 GTTCCATCCGGAACCTTTAGTTCCGACTTTCAAGTTTAAAAATTCGTTCTCCCAAACTT 16818
Qy 120 ----- 120
Db 16817 TAAATCGTACAGCAGAAATTTGATGTATGAAGACCAAGAGTGACACTTACAAATATA 16758
Qy 121 ----- SerGlyAsn 123
Db 16757 AAACCTAATATAACGCCCGGAAACAGTTATTTTCTAGTCGACTATCTTTTCAGGAAC 16698
Qy 124 AlaPheIleAspMetLeuAsnGlyAsnGlyIleProIleGlySerSerIleArgGlyLeu 143
Db 16697 GCATTCATCGATATGTTAAATGGAATGGAATTTCCAATTTGGAAGTAGTATTCGTGGATTA 16638
Qy 144 GluAspAlaIleArgThrGlnArgAspMetGluAsnThrAspProSerGlu----- 160
Db 16637 GAAGATGCTATTTCGGACACAAAGAGACATGGAGAACACTGATCCATCTGAGCAGGTAATT 16578
Qy 160 ----- 160
Db 16577 TCACTTTTAAGTTTGAATGACAGTCTATTGTGAAAAATATGGTAAATGTCATTTCATGTTA 16518
Qy 160 ----- 160
Db 16517 TTCTTGAAACAATAACATTTTTCGTCGAGTTTTCACCCGAGCTATATGAAGGTACCTA 16458
Qy 160 ----- 160
Db 16457 TGGCTCATTTTGAAGGAGAAAGCGTTCTACAGTCTATAATAAAGCATAGTTTACAAAAG 16398
Qy 160 ----- 160
Db 16397 GCCGGCGCGGCGGTAGAGAAATTTTTCGGGCCCGACCCGCGGCCGCCCTACCA 16338
Qy 160 ----- 160
Db 16337 CTGTTGAAAAATTTCCAATGAAAAATTTGGATTTTTTTTGAATAAAAAAGAAAAAGGT 16278
Qy 160 ----- 160
Db 16277 TACATTTTAAACAAAAATTAACAAAGTAAAAAGTAGTCAGAAATAGTCTGTTTTGTATAT 16218
Qy 160 ----- 160

D	b	15149	TTTGAATCTCTATTCCGAGCTAGCTAAATTTTTGGTTTCGTGCCTCTTTTAGAAGAACAATA	15089
Q	y	297	-----GlnAsnAlaGluValGlnSerLe	304
D	b	15089	CTACATATTAAACAAGGATTTCTAAATTTCCATAGTTTCAGAACGCCGAAGTCGAATCTCT	15030
Q	y	304	uLeuSerTyrGlnArgMetArgAspSerProIeuSerLysArgArgProLeuAlaMetAs	324
D	b	15029	TCTCTCTTACCCTCGAATGAGAGACTCTCCTCACTACGAAACGGAGGCCATTTGGCTATGAA	14970
Q	y	324	n-----	324
D	b	14969	TGGTTCGTCCTTTTTTTTGAAAAACGTTTTTAAGGATTTTTATGAAGTTTTTGAGAGAAGTTA	14910
Q	y	324	-----	324
D	b	14909	TACTTCTGCATTTTAAATTTTTTTTTTCAAATTTTAGGACTGCAGAGAGGATTTTCGATTTT	14850
Q	y	324	-----	324
D	b	14849	TATTGGGTATTATCATTTTAAAGTGGGGGACATGAIAAAAAAAAAATGTTCTCGAAATTC	14790
Q	y	324	-----	324
D	b	14789	GTTCAAAGATGCCCTCGAGGAGTTGAICTTGTGAAGCGAGGATTTAGGGTACGGCGCC	14730
Q	y	324	-----	324
D	b	14729	GCTACCGACTGAGCACCGCGGCAAGAGAAAGCGCACGCCGATTCATCATATAAAGGC	14670
Q	y	324	-----	324
D	b	14669	GAGGACGACAGATGTACTCGGAGAGTGACATTAAGCGTTGTACACCCTGACGGTGAAGATGA	14610
Q	y	324	-----	324
D	b	14609	TACTCTAGCTACAAAACAACTACGAAAAATTTCTGTCCCACCTTTAAGAAAAATCTT	14550
Q	y	324	-----	324
D	b	14549	GATTTGTTTCAATCTGAAAAATCGAAGTTTTTTTTTCAAGTAGCTTCAAGCTTTCTCC	14490
Q	y	324	-----	324
D	b	14489	TATATTCTCTTTGAAAATGACAGTCTGAACCTTTGTATTATATACTATTATCTAAATTTG	14430
Q	y	324	-----	324
D	b	14429	GATTTTCAAACCTTTCTAAAAACGGTAAAAATTTTGAAAAAACAACAAAGATTACGACACA	14370
Q	y	324	-----	324
D	b	14369	AAGAATACCAAGTCGGATTTCTTTCTAAAAATACGTCATGATTTTTCGTTTTTTT	14310
Q	y	324	-----	324
D	b	14309	AAACTAAATATTTTTGAATTAACAAATTTCTGAAGTTCAAATCTTTTGCATTTTTTCTAACAT	14250
Q	y	324	-----	324
D	b	14249	COGTTTCTCAACTTTTTTAGAAACCGTGAAAAACCTTTAAAAATGCTAACCACTGAGAGCA	14190
Q	y	325	-----AspGluAspGluSerAlaPheArgAlaMetGluAlaArgAlaLysLeuAspGl	342
D	b	14189	CITTTTCAGATGAAGACGAACAATTTTCGTGCGATGGAAGCCCGTGCMAAGCTTCGATCA	14130
Q	y	342	nLysSerGlnLeuValLeuGlyLeuHiGlyPhe-----GlyGluSerAspAspAs	359
D	b	14129	AAAAATCGCAATTAGTTCTTCGGATTTCATGGATTCGATGGGATGAAGATGAATGA	14070
Q	y	359	pGluAspGluAspGluAsnLeuIleAspProSerGluAsnSerPheArgArgAlaPr	379
D	b	14069	C---GATGAGGAAGATGAGAAATTCATTTGATCCATCAGCCAACCTCATTTCCGTGCTGCC	14013

Qy	379	oLeuArgLeuSerSerGlyPheValGluLysLeuLysSerAsnAspGluLeuLysSerAl	399	Qy	537	-----	537
Db	14012	ACTTCGATGTGCTCTCGATTGCGTTGAAACCTAAATCAATGATGAATGGAAGTGC	13953	Db	12933	TATTATTGCTTGCAAACTCATCGTTAAATTGGTACAAAGCTGTTTCTTCAACAATCTAATA	12874
Qy	399	aLeuAspArgLeuLysTyrrArgValAspValGluLysTyrrLeuAlaProLysProMe	419	Qy	537	-----	537
Db	13952	ATTGAACAGAAATCAAGTATCGAGTTGACGATGGGAAAAGTATTGCAACCAAGCCAAAT	13893	Db	12873	TCACAGGTTTCACCGTATCTTTTCATAAACTTGGTATCAGCAACTAACATTGAACGAACCT	12814
Qy	419	tGluPheAsnProLysProGlnProGlyTyrrPheAlaProArgLysIleProThrArgPr	439	Qy	537	-----	537
Db	13892	GGATTTCAATCCTAAACCCAGCCTCGATATTTGTACTCGAAAAATCCCACTAGACC	13833	Db	12813	CGCACTATCTGCAATTAATTAATTTTCGAAAACAGATAAGTTGAAACATGCTCAAAACTAA	12754
Qy	439	oArgLysMetLeuProLeuLeuIleGlySerAspProLysValGlnGluLuleArgAr	459	Qy	537	-----	537
Db	13832	TCGTAAATGCTTCGTTGTGTGATCGGATCAGATCCGGCACTCAAGAAGAAGTCAGAAG	13773	Db	12753	AAACCTATGTTCAAGTAAATTCCTCTTTCACAGTTAAATTTTGGTATTGCTAGCTGGG	12694
Qy	459	gHieProSerThrGluTrpLysIleAlaLysGluSerArgValLeuThrAsnLeuLysAs	479	Qy	537	-----	537
Db	13772	ATATCCAAAGCAAGATGGAAGATTGCAAAAGAAATCAAGAGTTCTGCAAAATTTGAAGAA	13713	Db	12693	AGTCATGTTGTGTGCTCTACATGCAACCGAAAAATTCAGTATTTCAGGTCAATCTCTTT	12634
Qy	479	nAsnProSerLeuAlaAlaLeuPheMetAspAspLysLeuGluAsnThrLeuLysGlyAr	499	Qy	537	-----	537
Db	13712	CAATCCAAGTCTTGCTGCACTTTTATGGATGAGAAATTCGAAAAATACGTTAAGAGGAAG	13653	Db	12633	CAAACCTACAAGAAAAAGTTTTCATTAATTTTAATATTTAAAAAACGAGATTTGGAGTT	12574
Qy	499	gGlnMetLeuThrAspGluGlnLysGlyArgThrArgValLysThrIleArgAlaLeuPr	519	Qy	537	-----	537
Db	13652	GCAATTTGCTGCTGATGAGCAGAGGGAAGAACAGAGTTTAAGACAAATCAGAGCATTTGCC	13593	Db	12573	CAGAAAAGTAGTGTAAATTCCAATAGGCGCATACAGGATGTCGATGTTTCATTGTTCTCGTC	12514
Qy	519	oArgLeuPheGlyAlaProThrAlaLysAlaGluMetIleAspAlaLysValPhe	537	Qy	537	-----	537
Db	13592	GAGATTTGTTGGACCAACTCGCAAGCGGAGATGATCGATGCAAAAGGTAAG-GAACA	13534	Db	12513	TGAAAAAGAAACCTTTTGAAGGAAATTTATTTAAATTCGCTTTCAAAAGCCTAGATGTTG	12454
Qy	537	-----	537	Qy	537	-----	537
Db	13533	GAAAAAATGAGTGTGTTTTTCGGAGTTTCTAGAACATGGAATAGAGCAGTTGAATAG	13474	Db	12453	TAAGCTTTTCATAAACTCGGTATCACCTACTTGAAATCGAATAAATTGAGCAATTAATAAA	12394
Qy	537	-----	537	Qy	537	-----	537
Db	13473	AACATGCATTAGTTTTTCAAAATGTAGTTGTGAACAGAAAAAACGTAATTTCAAAAT	13414	Db	12393	AGTAAGGCGCGAAGTTTGCATGTAGACACAAACCATGACTCCAGCTAGCAATAACC	12334
Qy	537	-----	537	Qy	537	-----	537
Db	13413	TGAACGTGTGCTTCTTCAATCATGCCGATCCAAAATACAAAAGACTGTAATTTCTTG	13354	Db	12333	AAAAATTAACGTGAAACAGGAAATTTACTTGAAACTGTGAAAAACAGGGAATTTACTTCA	12274
Qy	537	-----	537	Qy	537	-----	537
Db	13353	AGTAAAAAATAGTCGCCCCCATATTTAAACAACCTTCGTAAACCGACAGGAATATTTCAA	13294	Db	12273	AACATCATGAAAAATGAAAAACATTGCTTTCATAGTTTACCCCGTTTCACATAAATTTTGG	12214
Qy	537	-----	537	Qy	537	-----	537
Db	13293	TTTCACAATCAGTAGAATTTTCGGAATTTGTGTTCTAATTTTAAATTTTACAGCTGTTT	13234	Db	12213	TTATTGCTTGAAACCTTATCGTTAGTTCCGAATTGAAAAACAAAAATGCCAATTTTCA	12154
Qy	537	-----	537	Qy	538	-----GlnAspIleGluLysArgProIleProLeuPheGluPheGluProLysGlyAr	555
Db	13233	TAITTCGACTTTTTCAGAAAAATAGATTTCTAATGATGTTTGTAGAACTTTTGAGAAATCTA	13174	Db	12153	GGTTACTCAAGATATCGAAGACGACCGGTACCACCATTTGTTCTTCGAGCAAAAGGAAA	12094
Qy	537	-----	537	Qy	555	gHieThrArgLeuArgTrpThrGlyAlaAsnGluLysGluIleProGlyLeu	572
Db	13173	AAAAATCTAACAAAAATAAAAAATAAACTGATATACAACTATAGAAAGCCTCGATG	13114	Db	12093	ACACACTCGGTTGAGATGGACTCGAGCCAATGAGAAAGAAATTCGGGACT-CGGTGGGT	12035
Qy	537	-----	537	Qy	572	-----	572
Db	13113	TTGTAATAGTTGCTGTCATAAGCATGGTATCAACAATTGACATTPAAATAAATTTCCGACT	13054	Db	12034	TTCTTTTAAAAAATTGAATATGTAGAAAGTAGCTGAGAGAAGTATATTGAAAAACCAA	11975
Qy	537	-----	537	Qy	572	-----	572
Db	13053	TTTGGAGATATTCTGTTTCCGAACAGATAAAAAGGACCTAATTTTCCATTTTTCAGA	12994	Db	11974	ACAGAAGTATTTCAGTAGTATTTTGAAAAAGAGAAATTTTAAATGGTATACATTCAAAAC	11915
Qy	537	-----	537	Qy	572	-----	572
Db	12993	CATACTGAAAAATCGAAAAACCTCGGTAATAATATAATGTCCTGTTTTCACAACTAATTTT	12934	Db	11914	TACTAATTTTTCATCAACATCACTTCGTATCAACCCCATGAACCAAGTCTCGTTATTATA	11855
Qy	537	-----	537	Qy	572	-----	572

Db 11854 TGATTGTGAACGGGAAATTATAAATTCAGACATTTTGATTCACCCCTAATCTACAA 11795
 QY 573 -----G1 573
 Db 11794 TCCCTTTTGTCTTTCATTTGTGTAGTCCCTTATTTTCATTCAGCTATTTCCATATTCAGG 11735
 QY 573 ySerArgPheLeuLeuProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSe 593
 Db 11734 ATCTCGATTTCATCTCCCATCTCTCGACCCCAATGCGAGCTCTGAACACAGCTTTCTC 11675
 QY 593 rThrGlnGlyArgAlaAaGAspGluTTPAspThrMetPheLysIleProAsnAsnTTPAs 613
 Db 11674 AACTCAAGACGCTGCTCGATGATGGAATGGATACAAATGTTTCAAAATTCAAATTCAGGAA 11615
 QY 613 nProGlyAspGluValGlyPheLeuMetAsnSerLysThrLysArgPheValGlyGlyAs 633
 Db 11614 TCCAGAGAGAGAGTGGATTCGAATGAACTCAAAACCAACGATTTATTGGAGGAA 11555
 QY 633 nGlyAlaPheAspMetProAlaLeuGlyLeu 643
 Db 11554 CGGTGCATTTGATATGCCAGCGCTGGGTCTC 11524

RESULT 4

CBRG33D04/c 39089 bp DNA linear INV 04-NOV-2000
 LOCUS Caeenorhabditis briggsae cosmid G33D04, complete sequence.
 DEFINITION AC084550
 VERSION AC084550.1 GI:11095000
 KEYWORDS HTG.
 SOURCE Caeenorhabditis briggsae

ORGANISM

Caeenorhabditis briggsae
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditoidea; Rhabditidae; Peloderinae; Caeenorhabditis.
 1 (bases 1 to 39089)
 Washington University Genome Sequencing Center.

REFERENCE 2 (bases 1 to 39089)
 The C. briggsae Genome Sequencing Project

Unpublished
 Wilson, R.

Unpublished
 The sequence of C. briggsae cosmid G33D04

Unpublished
 3 (bases 1 to 39089)

Waterston, R.
 Direct Submission

Submitted (04-NOV-2000) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University,
 St. Louis, MO 63110, USA
 e-mail: jspieth@watson.wustl.edu

COMMENT

NOTICE: This sequence may not be the entire insert of this clone.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we provide a small overlap between
 neighboring submissions.

FEATURES

source

1. 39089
 /organism="Caeenorhabditis briggsae"
 /mol_type="genomic DNA"
 /strain="Gujarat G16"
 /db_xref="taxon:6238"
 /clone="G33D04"

ORIGIN

Alignment Scores:
 Pred. No.: 2,586-27 Length: 39089
 Score: 585.00 Matches: 141
 Percent Similarity: 41.37% Conservative: 10
 Best Local Similarity: 38.63% Mismatches: 11
 Query Match: 17.78% Indels: 203

DB: 3 Gaps: 2
 US-10-736-868-2 (1-643) x CBRG33D04 (1-39089)
 QY 1 MetIleLeuPheLeuPheLeuLeuLeuGlyPheCys----- 14
 Db 1805 ATGTTAATAATTTTACTAAATTTTCTGTTTCTCAATTTTGTAAAGTTTTTTACAGTTCAA 1746
 QY 15 -----Ileal 16
 Db 1745 AAGTTGGTCANTCGAAGGAAGAACTTTCGGTTACATCATTTCTTTAACTTTCTTAGA 1686
 QY 16 aProLeuSerAlaGlnSerProSerThrSerAspAlaProGlyAlaLeuLeuSerSerle 36
 Db 1685 TCCAATATCCCGGGCTCACCAACCAATTCAGATGTTCCCGGTGCTCTTTTATCTCTCT 1626
 QY 36 uValGlyLysSerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGluLeuMe 56
 Db 1625 TGCCACTAAAACCAACCAAACTTCCATTTAGCTCCATCCATGGAAGCACTAGAAATTGAT 1566
 QY 56 tGlyValGlnPheValAspAlaLeuIleLysLysGlyGlnMetGluMetAlaLysGlyVal 76
 Db 1565 GGGAGTTCAATTTGTGGATGCTTTAATTAAGGACAGATGGAAATGGCAAGAGGAGC 1506
 QY 76 aPheLysThrGlnLeuGluValLeuGluLysValHisProAspGlnPheAspLysTyL 96
 Db 1505 ATTCAAAACACACACTGGAGGTTCTGGAGAAAGTACATCCGGATCAATTTTGAGAAATATA 1446
 QY 96 sLysLeuLysValAspAspLeuAlaAlaAspAlaValMetGlnGlnAlaGluMetAlaL 116
 Db 1445 GAAACTGAAAGTGGAGGATTTAGCGGCAGATGTCAGTTATGCAACAGGAGCAAAATGTC 1386
 QY 116 sLeuGlnProLys----- 120
 Db 1385 GTTCGACCCGAAACTCGTAAGGAGAAATATAAAAGAGAGATCATTTACTACGAGCAATA 1326
 QY 120 ----- 120
 Db 1325 CGACTGAAATAATTAATTTCTAGTTAGCACATGCAACAGTAAAAAATTCAAAATTCAGTTT 1266
 QY 120 ----- 120
 Db 1265 GAGATAATGTGAGAAAAAAGTTTTTATTTTTCACATTTTAAATAATGTAAGTATCAATTTC 1206
 QY 120 ----- 120
 Db 1205 CGATCGAATCAATTTGGCAAAATCAATTAGAGGAATAAGTTTCTCGAAACTGAAAGATTTT 1146
 QY 120 ----- 120
 Db 1145 CAATGTGCAGATCTTGAAACTGAACTTTGCGAAAGTCTGAGAAACCTTTAAGTTAAACATT 1086
 QY 120 ----- 120
 Db 1085 GAACACTGTCTAATAGACAGGCGTCTTAACCTTCAAGTTTTCGAGACTTTTATATTTTGG 1026
 QY 120 ----- 120
 Db 1025 TGAAAAATGAGTAATAGTTCATCCCGGAACTTTAGTTCCGACTTTCAGTTTAAAAAATT 966
 QY 120 ----- 120
 Db 965 CGTCTTCCCAAACTTTAAATCGTACAAGCAGAAATTTGTATGTATGATGAGACCAAGAGTG 906
 QY 120 ----- 120
 Db 905 ACCTTACAATATATAAAAACTTAATATAACCGCCGAAACAGTTATTTTCTAGTCGACT 846
 QY 121 -----SerGlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIleProIleGlySe 138
 Db 845 ATCTTTTTCAGGAAACCGCATTCATCGATATGTTAAATGGAATTCGAATTCGAATTTGGAAG 786
 QY 138 rSerIleArgGlyLeuGluAspAlaIleArgThrGlnArgAspMetGluAsnThrAspPr 158

```

Db      785 TAGTATTCGTGGATTAGACATGCTATTCCGACACAAAGAGACATGGAGAACACTGATCC 726
Qy      158 oSerGluGlnIle 162
Db      725 ATCTGAGCAGGTA 713

RESULT 5
CQ575340
LOCUS   CQ575340
DEFINITION Sequence 3098 from Patent WO0171042.
ACCESSION CQ575340
VERSION   CQ575340.1 GI:41638873
KEYWORDS
SOURCE   Drosophila sp.
ORGANISM Drosophila sp.
          Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
          Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
          Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS Venter, J.C., Adams, M., Li, P.W. and Myers, E.W.
TITLE    Detection kits, such as nucleic acid arrays, for detecting the
          expression of 10,000 or more Drosophila genes and uses thereof
JOURNAL  Patent: WO 0171042-A 3098 27-SEP-2001;
          PE Corporation (NY) (US)
FEATURES
         source
         1. 3631
         /organism="Drosophila sp."
         /mol_type="unassigned DNA"
         /db_xref="taxon:7242"
ORIGIN

Alignment Scores:
Pred. No.: 0.0268 Length: 3631
Score: 181.00 Matches: 165
Percent Similarity: 35.51% Conservative: 102
Best Local Similarity: 21.94% Mismatches: 258
Query Match: 5.50% Indels: 228
DB: 6 Gaps: 37

US-10-736-868-2 (1-643) x CQ575340 (1-3631)
Qy      39 LysSerHisGlnIysLeuProLeu-----AlaProSer 49
Db      1031 AAGAACAAACAGAAAGATCCACTGATTGCGAGAGATTCCAAACCAAGAAAGTCCCTTAAG 1090
Qy      50 MetGluAlaLeuGluLeuMetGlyValGlnPheValAspAlaLeuIleLysLysGlyGln 69
Db      1091 AAGGAACAAACCAACCGGAGGA-----AAAGGCAAG 1123
Qy      70 MetGluMetAlaLysGlyAlaPheLysThrGlnLeuGluValLeuGluLysValHisPro 89
Db      1124 AAGGAGGAGCAAAACCGACTGTAAGAGAGTAAGAAAGAAAGAGAGAA-----1171
Qy      90 AspGlnPheAspLysTyrlsLysLeuLysValAspAspLeuAlaAlaAspAlaValMet 109
Db      1172 GACAGCAACAAAGTCAAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1231
Qy      110 GlnGln-----111
Db      1232 CAACAATTTGGATCAAGAAAGAAAGACTTCTGAACCAAGAGATGCATCGAATCCACACAA 1291
Qy      112 AlaGluMetAlaLysLeuGlnProLysSerGlyAsnAlaPheIleAspMetLeuAsnGly 131
Db      1292 TCGGTGACGGCCAAAAGAGATTCCAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1351
Qy      132 AsnGlyIleProIleGlySerSerIleArgGlyLeu-----143
Db      1352 AACGAATCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1411
Qy      144 -----GluAspAlaIleArgThrGlnArgAspMetGluAsnThrAsp---ProSer 159
Db      1412 TCCAGCAGCGAGCAGCAGGATGGATGGCGGATACGGACACCGCGGGCCACCGACAGACCCCTCC 1471

```

```

Qy      160 GluGlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeuProGlyLeuVal 179
Db      1472 GCCAA-----AAGGAGACTCGCTCGAAACAGGAAGAGTTCGCCGAA---GTA 1516
Qy      180 AlaAsnMetIleAlaGlyLysAsnPro-----PheLysMetProGlnGlnMetArgLys 197
Db      1517 ACCGGCGCCCGNAGGCTTACGCCCAACACAGCAACGACTCATGTCGAGCAAGGAAAG 1576
Qy      198 AlaGlnAlaAlaProSerValPheGlnGlnAlaLeuAlaGlnArgAlaMetLeuGly 217
Db      1577 GCTCGCGAGGAGAAG-----GAACAAAGCTCGCGAGGAGCGTTCGCTCAAG 1624
Qy      218 LysAsnAlaProValAlaGlyGlyArgGlyGluGlnArgMetMetAsnArgVal 237
Db      1625 CACAG-----GATAGGAGCACCGGAGCAGCAAAAGAAACAGAGCGC---1669
Qy      238 AspGlnArgMetGlnArgGluLeuGln---GluGluAspGluAspAspAspLeu 256
Db      1670 GACGAAAGAACAGCAGCGGAGCTGGAGGGATCAGAGGAGCAGCAGAGAGAGATG 1729
Qy      257 GluAspGluAspValProArgArgSerSerAspGlyGluProGlnSerGluAlaGlu 276
Db      1730 GAAAGGAGGAGAGAGGAGCGCAAGCGACAGCGAGGTGACTCCAAAGAACAGAGAGAG 1789
Qy      277 HisGlnArg-----ArgAspLeuAlaArgArgLeuLysSerProArgLeuLys 293
Db      1790 CGCAAGCGCAACGAGCGCAAGGAGGAGGTGCAACGCAAAAGAACGAGGAGCGTCGCAAG 1849
Qy      294 GluLeuLeuGlnAsnAlaGluValGln-----302
Db      1850 AAGGAGCAGGCGCTGAGGAGCGCGAGCAAAAGAAAGCGCGCGCGAGTCTTCTCC 1909
Qy      303 -----SerLeuLeu 305
Db      1910 AAGTTCTTCTGTCGCCAAGCAGCCAAAGTGCAGCGAGCGGATCCAAACACACATCTCTACCTG 1969
Qy      306 SerTyrGlnArgMetArg---AspSerProLeuSerLysArgArgProLeuAlaMetAsn 324
Db      1970 GAGCAGGCAAAAGCAGCTGCGACAGCTCCAGCGCAGCAGCAGCAGCAGCAGCAGCTG 2020
Qy      325 AspGluAspGluSerAlaPheArgAlaMetGluAlaArgAlaLysLeu-----340
Db      2021 -----GCTTTCCAGCATTCAGATCAAGGAGCAGCATGCTGTAGCAGCCG 2065
Qy      341 -----AspGlnLysSerGlnLeu---ValLeuGlyLeuHis 351
Db      2066 ATCGTCGCAACTCGCTCGGCCAGGAGCAGCGCTCCAGCTAGATGGACTTTTTCGACAC 2125
Qy      352 GlyPheGlyGluSerAspAspGluAspGluGluAspGluAsnLeuIleAspProSer 371
Db      2126 CGGAGCAGGAAGCTCAGCAGCAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2176
Qy      372 GluAsnSerPheArgArgAlaPro-----LeuArgLeuSer 383
Db      2177 -----AGACGAAAGCGCCGAATCGGCTAAATCTCTATCTTAGTGAGTTGAGC 2224
Qy      384 SerGlyPheValGluLysLysLysSerAsnAspGluLeuLysSerAlaLeuAspArgIle 403
Db      2225 AGCGGACGGCGCAAGCGCTCAAGATGCAGCGGATGTGAAG---CTCCAAAGACGACCC 2281
Qy      404 LysTyrArgValAspAsp-----ValGluLysTyrLeuAlaPro-----416
Db      2282 AAGCAGGAGGAGATGATGACGATGTTCAGTGATAGATTACCTTTTCCCGCGGGGTCTG 2341
Qy      417 -----LysProMetGluPhe 421
Db      2342 CCCATCGAGGTGGAGCAGCCCAAGCAGCTGACCCGAATGAAGCCCAAGATCTGCAATTC 2401
Qy      422 AsnProLysProGlnProGlyTyrPhe-----Ala 431
Db      2402 GCCGACAAACCGGAGGCCACCGTACTACGGCACCTGGCGTAAAAAGAGCAGCAGCATCTCC 2461

```


3427	Db	GAGCCGTTAAGGACCTTCATCGAGCAGCCTTGATCGCCCAAGGAGGCCCAACAG-----	3477
47	Qy	AlaProSerMetGluAlaLeuGluLeuMetGlyValGlnPheValaAspAlaLeuIleLys 66 ::: :::	
3478	Db	-----GTCAC 3483	
67	Qy	LysGlyGlnMetGlu-----MetAlaLysGlyAlaPheLysThr---GlnLeuGluVal 83 ::: :::	
3484	Db	AAGGACGAGGTGGCGTTCGGTGTCAACAAGATCTTCTTCGCTCGGCGCCAGCTGGCGCC 3543 ::: :::	
84	Qy	LeuGluLysValHisProAspGlnPheAspLysTyrlsLysLysLeuLysValaAspAspLeu 103 ::: ::: ::: ::: ::: :::	
3544	Db	ATCGAGGAGCTCGCGAGCAGGCCATCTCCAAG-----ATGGTGGTGTGCATC 3591 ::: :::	
104	Qy	AlaAla-----AspAlaValMetGlnGlnAla 112 ::: :::	
3592	Db	CAGGCGCGCGCGCGCTTCCTCGCGCGCGCATGTACGACAAAGATGCGCGAGCAGACC 3651 ::: :::	
113	Qy	GluMetAlaLysLeuGlnProLysSerGlyAsnAlaPheIleAspMetLeuAsnGlyAsn 132 ::: :::	
3652	Db	GTGTCGCCAAGATCTCGAGCGCAACATCCGCGCTCGGTTCGAGTCAAGAACTGGGCC 3711 ::: :::	
133	Qy	GlyLeuProIleGlySerSerIleArgGlyLeu----- 143 ::: :::	
3712	Db	TGTTACCATCTTACGTCAAGGCGCCCGCTCATCTCGCAGCGCAACTTCCAGAGGAG 3771 ::: :::	
144	Qy	--GluAspAlaIleArgThrGlnArgAspMetGlu----- 154 ::: :::	
3772	Db	ATCGAGACCTCAAGAAGCAGGTCAAGGACCTCGAGAAGGAGTTCGCGCGCTCAAGGAC 3831 ::: :::	
155	Qy	-----AsnThrAspProSerGluGlnIleAlaLysAlaValMetAspLysPheGln 171 ::: :::	
3832	Db	GCCAAAGCGCAAGCTCGACAGGAGAGCAGCTCGCGGAGGAGACGCCGCAAGCTCGAG 3891 ::: :::	
172	Qy	ThrGlnIle-----LeuProGlyLeuValAla----- 180 ::: :::	
3892	Db	AAGGACCTCGCGCGCTCAAGCTCAAGATCTCTCGATCTCGAGGCGAAGAGCGCGATCTC 3951 ::: :::	
181	Qy	-----AsnMetIleAlaGlyLysAsnProPheLysMetProGlnGlnMetArgLys 197 ::: :::	
3952	Db	GAGGAGGACACCGCGTCTGCAGAAAGAGTGGCGCGCTCGAGGAGGAGCTGCGAGGAG 4011 ::: :::	
198	Qy	AlaGlnAlaAlaProSerSerValPheGlnGlnAlaLeuAlaGlnArgAlaMetLeuGly 217 ::: :::	
4012	Db	GAGACCTCGCGCTCCAAACGACATCTCTGGAGCAG-----AAGCGCAAGCTCGAGGCC 4062 ::: :::	
218	Qy	LysAsnAlaProValAlaGlyArgGlyArgGlyGluGluGlnArg-----MetMet 233 ::: :::	
4063	Db	GAGAAGGGCGAGCTCAAGCGCTCGCTCGAGGAGGAGGAGCGCAACCGCAGGCGCTCCAG 4122 ::: :::	
234	Qy	MetAsnArgValAspGlnArgMetGlnGlnArgGluGluGlnGluAspGluAspAsp 253 ::: :::	
4123	Db	GAGGCCAAGACCCAGTCCAGTCCGAGCGCAACGAGTTCGAGGACAAAGTACGAGGACGAG 4182 ::: :::	
254	Qy	-----AspAspLeu-----GluAspGluAspValProArgArgArgSerSer 267 ::: :::	
4183	Db	GCGCCGCGCAGCAGCTCGCTCAAGAAGAAGGAGGAGGAGCCTGTTCGCGCGAGCTCCCGCAG 4242 ::: :::	
268	Qy	AspGlyGluProGlnSerGluAlaGluHisGlnArgArgAspLeuAlaArgArgLeuLys 287 ::: :::	
4243	Db	ACCAAGGAGCGCTGGCGCCGCGAGAACTCTCCGAGACCCCTCGGCTCCAGGCTCAAG 4302 ::: :::	
288	Qy	SerSerProArgLeuLysGluLeuLeuGlnAsn-----AlaGluVal 301 ::: :::	
4303	Db	AACCCAGCGCGCGCGCCGAGCAGCTGGCGCAACAGAGCTCGACGACGCTACCGGCCCAACAG 4362 ::: :::	
302	Qy	GlnSerLeuLeuSerTyrlGlnArgMetArgAspSerProLeuSerLysArgArgProLeu 321 ::: :::	
4363	Db	CTCCAGCTTTGAGAGACCAAGAAGAGCGCTCGAGGAGGAGCTCGCCAGACCCGC----- 4416 ::: :::	
322	Qy	AlaMetAsnAspGluAspGluSerAlaPheArgAlaMetGluAlaArgAlaLysLeuAsp 341 ::: :::	
4417	Db	GCCCAAGCTCGAGGAGAGAGAGCGCAAGGAGCGCCGAGGCTTCAAGGCCAAG----- 4470 ::: :::	

```
Qy 342 GlnLysSerGlnLeuValLeuGlyLeuHisGlyPheGlyGluSerAspAspGluAsp 361
Db 4471 -----CAGCTC-----
Qy 362 GluGluAspGluLeu-----IleAspProSerGluAenSerPheArgAla 378
Db 4501 TCCGAGTGCAGCTCGCTCAAGAGCAAGCTTCGGCCCGCAGGAAGTCGCTCAAGACCGCC 4560
Qy 379 ProLeuArgLeuSerGlyPheValGluLysLeuLysSerAsnAspGluLeuLysSer 398
Db 4561 AAGGACCAG--AACCAGCGCTCGAGCAGCTTCGAGGACGAGCGCCACCGTGGCGGCC 4617
Qy 399 AlaLeuAspArg-----IleLysTyrArgValAspAspValGluLysTyrLeu 414
Db 4618 AACGTCGACCAAGCAAGAGCGCTCGAGCGCAAGCTCCGAGCTCGAGGACCGAGTC 4677
Qy 415 Ala-ProLysProMetGluPheAsnProLysProGlnProGlyTyrPheAlaProArgLy 434
Db 4678 ACCGCTCGAGCGCCAGAGACGCGCGCCG-----CCAGGCC 4719
Qy 434 sileProThrArgProArgLysMetLeuProLeuLeuIleGlySerAspProLysValG 454
Db 4720 AAGACCTCAAGACCCAGTTCGAGACCAAGCGCGCTCGAGGAGCGCGAGCTCG 4779
Qy 454 nGluGluLeuArgArgHis-----ProSerThrGluTrpLysIleAlaLysG 470
Db 4780 CGCGCGCGCTCGAGAGAGGAGCGCAAGACCGCTCGACGAGTGG-----4825
Qy 470 uSerArgValLeuThrAsnLeuLysAsnAsnProSerLeuAlaAlaLeuPheMetAspAs 490
Db 4826 -----CCAG-CTCACC-----GACCTCGAGCG 4850
Qy 490 pLysLeuGluAsnThrLeuLysGlyArgGlnMetLeuThrAspGluGlnLysGlyArgTh 510
Db 4851 CGAGCGCGACTCGGGCGCCAGAGAGCGAGCGCAAGCTC-----AACAC 4892
Qy 510 rArgValLysThrIleArgAla-----LeuProArgLeuPheGlyAlaProTh 526
Db 4893 CCGCATCAGCGAGCTCGACTCGAGCTCGAGACCGCGCCCAAGCGCGCGCTCGTC 4952
Qy 526 rAlaLysAlaGluMetIleAspAlaLysValPheGlnAspIleGluGluArgProfile-P 546
Db 4953 CGAGGAGTGAAGCGCTCGAGGGCGAGCTC---GAGCGCTCGAGGAGAGCTCTCAC 5009
Qy 546 roProLeuPhePheGluProLysGlyArgHisThrArgLeuArgTrpThrGlyAlaAsnG 566
Db 5010 CGCCCGAGGCGCGCGCGCGCGCGAGAGAACTCGACAAGCCCAACCTCGAGCTCGA 5069
Qy 566 lLysGluIleProGlyLeuGlySerArgPheIleLeuProSerLeuAspProThrMetP 586
Db 5070 GGAGC-----TCGCGCAGGAGCGCGAGCG 5096
Qy 586 roAlaLeuAsnThrAlaPheSerThrGlnGlyArgAla 598
Db 5097 CGCGCGCGACACGACAGCTCGTCAAGGACACCGCA 5134

RESULT 8
SPAC29E6/c
LOCUS SPAC29E6 19514 bp DNA linear PLN 18-OCT-1999
DEFINITION S.pombe chromosome I cosmid c29E6.
ACCESSION Z66525
VERSION Z66525.1 GI:1044926
KEYWORDS coiled-coil; cysteinyl-tRNA synthetase; G-beta repeat; p115;
peptide methionine sulfoxide reductase; tdf1; TFIID; transcription
initiation factor iid; trp-asp repeat; U4/U6 small nuclear
ribonucleoprotein; U4/U6 snRNP; vesicular transport factor; WD
domain.
SOURCE Schizosaccharomyces pombe (fission yeast)
ORGANISM Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomycetes.
```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 19514)
Jones, L., Murphy, L., McNeill, A., Simpson, I., Harris, D.,
Barrell, B.G., Rajandream, M.A. and Walsh, S.V.
Direct Submission
Submitted (23-OCT-1995) Schizosaccharomyces pombe chromosome I
sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge
CB10 1RQ E-mail: barrell@sanger.ac.uk
Notes:
Details of yeast sequencing at the Sanger Centre are available on
the World Wide Web.
(URL, <http://www.sanger.ac.uk/Projects/S.pombe/>)
Protein coding regions (CDS) have been predicted with the help of
computer analysis using the Genefinder program in PomBase (an ACEDB
database) with additional predictions for the branch-acceptor sites
supplied by the program Sp3splice. CAUTION: It is possible that for
any individual CDS we may have underestimated or overestimated the
number of introns/exons or we may not have chosen the correct
splice donor/acceptor sites. CDS are numbered using the following
system eg SPACSH10.01c. SP (S. pombe), A (chromosome 1), c5H10
(cosmid name), 01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous. The length
in codons is given for each CDS.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions.

FEATURES
source

1. .19514
/organism="Schizosaccharomyces pombe"
/mol_type="genomic DNA"
/strain="972h-"
/db_xref="taxon:4896"
/chromosome="I"
/map="IR"
/clone="cosmid c29E6"
1. .804
/genes="SPAC29E6.01"
1. .804
/genes="SPAC29E6.01"
/notes="SPAC29E6.01", len:267, SIMILARITY:Caenorhabditis
elegans.. YSS1 CAEBL, hypothetical 80.3 kd trp-asp repeats
containing protein k10b2-1 in chromosome ii., (701 aa),
fasta scores: opt: 686, E():0, (38.7% identity in 269 aa)"
/codon_start=1
/label="SPAC29E6.01
/product="hypothetical trp-asp repeat-containing protein"
/protein_id="CAA91423.1"
/db_xref="GI:1044927"
/db_xref="UniProt/Swiss-Prot:Q09855"
/translations="DRIVSVMDVNSRFILYKLYGHSGSVLCLDFCRRNLLVSGSDS
TIIIDWQNRRLPKVYFGHTDNLGVVSVSENYIISSRDHTARWRDLATSPACMH
VLRGHLASVNSVQYSKTLGLIVTASDRLTRWDITTHGICIRIIHAHQIGACQYNG
KFIVSGSDLTIRIFEASSGKLRLMLQGHEDLIRTVRFNDEKIVSGYDGTIRWNFN
TGEQHCVLHNSRNSRVFGLQFDHRRIIIACTHSSEILVWNPDDCLDCTFF"
40. .150
/genes="SPAC29E6.01"
/note="Match to PF00400 WD40, WD domain, G-beta repeat
Score 42.94"
163. .267
/genes="SPAC29E6.01"
/note="Match to PF00400 WD40, WD domain, G-beta repeat
Score 34.79"
298. .408
/genes="SPAC29E6.01"
/note="Match to PF00400 WD40, WD domain, G-beta repeat
Score 42.65"
419. .525
/genes="SPAC29E6.01"
/note="Match to PF00400 WD40, WD domain, G-beta repeat
Score 20.89"
538. .648
/genes="SPAC29E6.01"

```

/note="Match to PF00400 WD40, WD domain, G-beta repeat
Score 48.28"
/misc_feature
/gene="SPAC29B6.01"
/note="PS00678 Beta-transducin family Trp-Asp repeats
signa ture"
gene
join(1228..1308,1359..1517,1564..2223,2266..2576,
2660..3077)
/gene="SPAC29B6.02"
/note="SPAC29B6.02"
CDS
join(1228..1308,1359..1517,1564..2223,2266..2576,
2660..3077)
/gene="SPAC29B6.02"
/note="SPAC29B6.02, len:542, SIMILARITY:Homo sapiens,
O43446, u4/u6 small nuclear ribonucleoprotein hprp3, (683
aa), fasta scores: opt: 946, E():0, (35.2% identity in 542
aa)"
/codon_start=1
/label=SPAC29B6.02
/product="putative u4/u6 small nuclear ribonucleoprotein"
/protein_id="CAA91424.1"
/db_xref="GI:1044928"
/db_xref="UniProt/Swiss-Prot:Q09856"
/translation="MDNRKRRLSSQLSESSPNQKPNAMERIKRRLQELQRLAQ
QAQVPEKRSNGNAGNETQNRISLSEKTAQFNANIPKSGLFRDDNGAGGL
KVGHPVLDDNGIQTILTPENRKRTASFTKGVSLSHQILKPPAITEQNPFLDTAP
TRLEDSFDPYPTKESKTRGSRNLHNSGKFTIEANQARQARLEDKRLIALHS
HKAGTEDELDITSKISGRDTPNIEWMDLPFKDYNDGNNWLIDGQSIINSAIQ
LPIVLPVAKNPSSHVLTKEQKIRQTRAARAKEKQDRLQGLIEPPEPKVK
LSNLMHVGDDAIKDPTKIEAVKQVEERLHRERNEERKLTPERKEKAPRKDE
DSAGRLCLVFRIKYLAHRPRLKIDLNKQWATGVCILNANFLNLFVEAGQAIKK
LSRMLERDWDTSRNSITAGNKLVDTEGRELNVTENTCNLVWEGEIGRRAPRYWS
FKPCSENDASKYLEEQGAEHFWMLAKWSENV"
/misc_feature
1309..1314
/gene="SPAC29B6.02"
/note="gtatgt, splice donor sequence"
/misc_feature
1345..1358
/note="gtatgt, splice donor sequence"
/misc_feature
1518..1523
/gene="SPAC29B6.02"
/note="ctaatcattatag, splice branch and acceptor"
/misc_feature
1550..1563
/note="gtatgt, splice donor sequence"
/misc_feature
2224..2229
/gene="SPAC29B6.02"
/note="ctaatcattttag, splice branch and acceptor"
/misc_feature
2251..2265
/note="gtaat, splice donor sequence"
/misc_feature
2393..2397
/note="ctgattattttag, splice branch and acceptor"
/misc_feature
2577..2582
/note="sequence duplicated by E. coli transposon tn1000,
CA UTION: tn1000 (gamma delta) transposon sequence and
duplicat ed 5bp sequence deleted from this sequence"
/misc_feature
2647..2659
/gene="SPAC29B6.02"
/note="gtatgt, splice donor sequence"
/misc_feature
2677..2683
/gene="SPAC29B6.02"
/note="ctaatcattatag, splice branch and acceptor"
gene
complement(join(3131..4432,4588..6237,6312..6479))
/gene="SPAC29B6.03c"
CDS
complement(join(3131..4432,4588..6237,6312..6479))
/note="SPAC29B6.03c, len:1039, SIMILARITY:Bos taurus,
P115 BOVIN, general vesicular transport factor p115, (961
aa), fasta scores: opt: 608, E():1.2e-24, (23.3% identity
in 980 aa)"
/codon_start=1
/label=SPAC29B6.03c
/product="putative vesicular transport factor"
/protein_id="CAA91425.1"

```

```

/db_xref="GI:1044929"
/db_xref="UniProt/Swiss-Prot:Q09857"
/translation="MDIPKSYNVILSPPKVQQADETISKCLDRLEHATLPEDRKA
LGICSPAREFKELVAAGLKIQSLRQDYPDELLKVTILETCLILTRHDDSRASDT
GLWADQFILNQDNIQCLLOSHKDFVRLYSVELFSAIILSCRETEKDCQCLTPPSA
ISSIMVPLRDSIEFVRNADLYFLSELVKDCTSIQKLVVFENAFELFSLLENGVDG
GIIALEALKFNLVLLKONISNQNYFRESNHPSLIKLISVDTFVDTWDTSRVQCIVIE
ALYALQSLVPIGLSSIANONAVSNHGIDVLLLTATHPDLLFPDQKISMTLAAHV
YSNARQNSFVDSFTFDDIKNTDVLTCFLPFLASISPSHRYSVAFKISRLTSENDEL
SSKFLKIIHAYTHQDNRLNIQGYLDLVHLSDDQDNDNFTSTILTYLVINDQDK
YLLCSIFLQDNDNDESDSEKDTFFQCVSTKLIATRHENALQNGVGYTLTILALV
YGNPDSVKDFLSESIQTLFTALMDESSANSYIQGMIAVFLSLVYVYIESPSVK
SDVYNTAISAKRVDFNRLQRLRMMLYEITFLSKQMKQLKSLREIDNTKEALDL
SVKRSIQEELKLSLTKTNLEOTQLAKEHEELDNQOKLVDLRIELDYTKSNC
KOMEEMQVLRGEHSEIKDFIESHSLTKOLDIDDKQPGIISKNRDLLESEKSKS
LANSIALESKNKLENDLNLLEKLNKNADTESFKNTIRBAELSKALNDNLGNKE
NIISDLKNKJSESTRLOQSLQDNQKQIETLNERISAADELSSMESINKNOANE
LKAQKCSNLQEKINFGNKLAKHEHTEKISLEKLELAATATSTLSKELTKVSEND
SLKSVNDNDQNEKSVNNEKFKVSOALAEANKEKNARDEIERLKVDIIIGLQNASLN
MOSLKDSDNRTISDLESKNKELEKLEADEYWLIVELLESKRTKDKELLRQCGQAV
SEDEQSSEE"
/misc_feature
complement(4433..4449)
/gene="SPAC29B6.03c"
/note="ttgacgaagcacaaag, splice branch and acceptor"
/misc_feature
complement(4582..4587)
/gene="SPAC29B6.03c"
/note="gtatgt, splice donor sequence"
/misc_feature
complement(6238..6250)
/gene="SPAC29B6.03c"
/note="ctacaaatttag, splice branch and acceptor"
/misc_feature
complement(6306..6311)
/gene="SPAC29B6.03c"
/note="gtatga, splice donor sequence"
6914..7351
/gene="SPAC29B6.04"
6914..7351
/gene="SPAC29B6.04"
/note="SPAC29B6.04, len:145, SIMILARITY:Saccharomyces
cerevisiae, NNFL_YEAST, nnf1 protein, (201 aa), fasta

```

```

Alignment Scores:
Pred. No.: 0.305 Length: 19514
Score: 179.50 Matches: 131
Percent Similarity: 37.34% Conservative: 124
Best Local Similarity: 19.18% Mismatches: 289
Query Match: 5.45% Indels: 139
DB: 8 Gaps: 28

US-10-736-868-2 (1-643) x SPAC29B6 (1-19514)
Qy 17 ProLeuSerAlaGlnSerProSerThrSerAspAlaProGlyAlaLeuLeuSerSerLeu 36
Db 18056 CCTTTTCAGCCTCTTATCCCTCTTAAAGCAGATGAGAAAATAGTATCATCTTCGTATCTT 17997
Qy 37 ValGlyLysSerHisGlnLysLeuPro-----LeuAlaProSerMetGluAla 52
Db 17996 TATTAACCTCGGAGTAGTTTAAACGCTGAAGGAGGTATCTTAAACGCTAGCGATGACCTA 17937
Qy 53 LeuGluLeuMetGlyValGlnPheValAspAlaLeuLeuLysLysGlyGlnMetGluMet 72
Db 17936 TTGAAGAATGATGGAGAAGTTTATAGAATGATGGAACAGTTCGTAGCTGACGCTGATG 17877
Qy 73 Alalys-----GlyAlaPheLysThrGlnLeuGluVal 83
Db 17876 CAGCGTGAAGATAATACAAATTTTCACGAACACGAGATTATACGAAGCGGTTGGAATAT 17817
Qy 84 LeuGlyLysValHisProAspGlnPheAspLysTyrLysLysLeuLysValAsp----- 101
Db 17816 GACGAAGATGAGAGAGAGAGAGAGAT-----GTAGATGAGGAC 17775
Qy 102 -----AspLeuAlaAlaAspAlaValMetGlnGlnAlaGluMetAlaLysLeuGlnPro 119

```

Db 17774 GAGTAGATCTAATGACCGATGAACAAGGATGAAGAGGACCGACGATGTTCCAAATT 17715
Qy 120 LysSerGlyAenAlaPheIleAspMetLeuAsnGlyAenGlyIleProIleGlySer 139
Db 17714 TTTGCTGCCGCTTATTGAGCAACGCTGTG----- 17685
Qy 140 IleArgGlyLeuGluAspAlaIleArgThrGlnArg----- 151
Db 17684 TTGCAAGCTTATCGTGAGAAAGTTGCTCAACACGCGCAAGCTAAACCTTTTAGAAGAGATT 17625
Qy 152 AspMetGluAenThrAspProSerGluGlnIleAlaIleAlaValMetAspLysPheGln 171
Db 17624 GAGGAGAGAACCAACCAACAGACGCTGAGCTCAAAAAATAAGAGAGAGAGAGAAA 17565
Qy 172 ThrGlnIleLeuProGlyLeuValAlaAsnMetIleAlaGlyIleAsnProPheLysMet 191
Db 17564 AAGAGG-----GATAAAAAAGAGCAATTGAAATTG 17535
Qy 192 ProGlnGlnMetArgLysAlaGlnAlaAlaProSerSerValPheGlnGlnAlaLeuAla 211
Db 17534 GCGAAGGAGGAGAGCGCTCAAGACGCGAAGCTGAGCGTCTTGCTGAACAGGCT---GCT 17478
Qy 212 GlnArgAlaMetLeuGlyLysAsnAlaProValAlaGlyArgGlyGluGlnArg 231
Db 17477 CAAAAAGCATGGAGCTAAAGACAGACAGAAAGCTCGCAAGAAACCGGAGAGCAACGT 17418
Qy 232 MetMetAenArgValAspGlnArgMetGlnGln-----ArgGluLeuGlnGlu 248
Db 17417 -----TTGAAGCGTGAGCAGGAGAGAAAAAGCAGCAAGAACTTGAGCGCTCAGAAAAAGAGAG 17364
Qy 249 GluAspGluAspAspAspLeuGluAspGluAspValProArgArgArgSerSerAsp 268
Db 17363 GAAAGCAAAAGCAGAGAGGAGAGAGAGAACTGAAGAGCAACCAAGAGCGGTGAC 17304
Qy 269 GlyGluProGlnSerGluAlaGluHisGlnArgArgAspLeuAlaArgArgLeuLysSer 288
Db 17303 CGTGAANAATGGCGAGAGACACGACTAAGAGAGAGAGAAACGTAATTTTGGAA 17244
Qy 289 SerProArgLeuLysGluLeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrGln 308
Db 17243 GAAAGGAAGCGAAGGAAAAATTAGACAAAGAGAGGAA----- 17202
Qy 309 ArgMetArgAspSerProLeuSerLysArgArgProLeuAlaMetAsnAspGluAspGlu 328
Db 17201 -----CGTCGCGCTGCTGAGTTACTTGAAAAAGAAATCAGAAAG 17163
Qy 329 SerAlaPheArgAlaMetGluAlaArg---AlaLysLeuAspGlnLysSerGlnLeuVal 347
Db 17162 AAAGAAAGGAGATTAAAGAGAGCTAAGATCGCTCTTCTTTCACCTAATCAGACTAAA 17103
Qy 348 LeuGlyLeuHisGlyPhe----- 355
Db 17102 GAGGTTCTGATGCTGTGACTACTTCTTCAACTTGGCCCTTTTGAAGAAAAAGCGCAT 17043
Qy 356 SerAspAspAspGluAspGluAspGluAspGluAsnLeuIleAspProSerGluAsnSerPhe 375
Db 17042 CTCGTAAACGATGAGTAAAGCTTTCTTCTCACTCTCTAGATTCTGTACCTTAATGCTTTG 16983
Qy 376 ArgArgAlaProLeuArgLeu-----SerSerGlyPheValGluLysLeuLysSerAsn 393
Db 16982 CGGCAAGCTCCATAGGCTTAAAGAACACAAATTAATTTTCGGAGCGCAATGCTTCTTCT 16923
Qy 394 AspGluLeuLysSerAlaLeuAspArgIleLysTyrArgValAspAspValGluLysTyr 413
Db 16922 AATCGCTTAATCTCTGTTATTTTCACTTTTAAATCCGTCACCCCTTTAAATTTCT--- 16866
Qy 414 LeuAlaProLysProMet-----GluPheAsn 422
Db 16865 CTTGAGGCTTAATCTTTAAATAGAACACTTAATTAATTCGGTAAACTGATTTTGGT 16806
Qy 423 ProLysProGlnProGlyTyrPheAlaProArgLysIleProThrArgProArgLysMet 442
Db 16805 AGAAGGCCAAT---GGTTTACATCTCCATCTTCACTCTCACTTAAGTAATTCACCAAT--- 16752

Qy 443 LeuProLeuLeuIleGlySerAspProLysValGlnGluGluIleArgArg-----His 460
Db 16751 -----TTTGATTAAATCCAAATGCTCGCATTTCTCTTAGTAGAGCAAAATTCA 16704
Qy 461 ProSerThrGluTrpLysIleAla-----LysGluSerArgValLeu 474
Db 16703 CCGGTGCATCATTTATCTTTTTCACACCTCTCTTCAAGCTGCTTAATAATATTCATTG 16644
Qy 475 ThrAsnLeuLysAsnAsnProSerLeuAlaAlaLeuPheMetAspAspLysLeuGluAsn 494
Db 16643 AATAAGGAGCAACACGCTACAGCTTGCTTAATTCATTTTCAAGTCCACAAATTAAGTCCT 16584
Qy 495 ThrLeuLysGlyArgGlnMetLeuThrAspGluGlnLysGlyArgThrArgValLysThr 514
Db 16583 CTAGTA-----AATAGAGTTTGAACGAGCGCTCTAGTTCCCTTTATCGTCTCTCAAGT 16530
Qy 515 IleArgAlaLeuProArgLeuPheGlyAlaProThrAlaLysAlaGluMetIleAspAla 534
Db 16529 TTGAAA-----TCCCCCTTATCCAAAGAGCGCTCTTAATCA 16491
Qy 535 LysValPheGlnAspIleGlu-----GluArgProIleProPro 547
Db 16490 CAAGACATGACGAGTATAACTTCTCACTGTCACTATATCCGAAATAAGCTATCTCTCT 16431
Qy 548 LeuPhePheGluProLysGlyArgHisThrArgLeuArgTrpThrGlyAlaAsnGluLys 567
Db 16430 ATTTGTGCTGCTAGTCAAGGAGTCTCTCAAGCTGAAGAAATAATTTGTCCAACACAGAG 16371
Qy 568 GluIleProGlyLeuGlySerArgPheIleLeu---ProSerLeuAspProThrMetPro 586
Db 16370 GAA-----AGGATGGGTAGTGTGACATTTAGATGACAAAACAGATTTCTGTATCACA 16317
Qy 587 AlaLeuAsnThrAlaPheSerThrGlnGlyArgAlaArgAspGluTrpAspThrMetPhe 606
Db 16316 GCTTCTTAATCA-----ACCACCTTCTGCTGTCTCAGAGATGAATCTAGTAAT----- 16269
Qy 607 LysIleProAsnAsnTrp-----AsnProGlyAspGluValGlyPheLysMetAsn 623
Db 16268 -----CCTAATAATTATGATGTTGTTAAACGCGCTTAAACAAACACCTCGAAATAATCT 16215
Qy 624 SerLysThr 626
Db 16214 CGTCCGACT 16206
RESULT 9
SPAC30/c 36315 bp DNA linear PLN 20-JUN-2003
LOCUS S.pombe chromosome I cosmid c30.
DEFINITION AL136538
ACCESSION AL136538.1 GI:6723883
VERSION
KEYWORDS
SOURCE
ORGANISM
Schizosaccharomyces pombe (fission yeast)
Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomycetes.
REFERENCE
1 (bases 1 to 36315)
AUTHORS
Wood, V., Gwilliam, R., Rajandream, M.A., Lyne, M., Lyne, R.,
Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., Basham, D.,
Bowman, S., Brooks, K., Brown, D., Brown, S., Chillingworth, T.,
Churcher, C., Collins, M., Connor, R., Cronin, A., Davis, P.,
Fitzell, T., Fraser, A., Gentles, S., Goble, A., Hamlin, N., Harris, D.,
Hidalgo, J., Hodgson, G., Holtroyd, S., Hornsby, T., Howarth, S.,
Huckle, E.J., Hunt, S., Jags, K., James, K., Jones, L., Jones, M.,
Leather, S., McDonald, S., McLean, J., Mooney, P., Moule, S.,
Mungall, K., Murphy, L., Niblett, D., Odell, C., Oliver, K., O'Neill, S.,
Pearson, D., Quail, M.A., Rabinowitsch, E., Rutherford, K., Rutter, S.,
Saunders, D., Seeger, K., Sharp, S., Skelton, J., Taylor, R.G.,
Tivey, A., Walsh, S., Warren, T., Whitehead, S., Woodward, J.,
Voickaert, G., Aert, R., Robben, J., Grymonprez, B., Weijens, I.,
Vanstreels, E., Rieger, M., Schafer, M., Muller-Auer, S., Gabel, C.,

Fuchs, M., Dusterhofs, A., Fritze, C., Holzer, E., Moestl, D., Hilbert, H., Borzym, K., Langer, I., Beck, A., Lehmach, H., Reinhardt, R., Pohl, T. M., Eger, P., Zimmermann, W., Wedler, H., Wambutt, R., Purnelle, B., Goffeau, A., Cadieu, E., Dreano, S., Gloux, S., Lelaure, V., Mottier, S., Galibert, F., Aves, S. J., Xiang, Z., Hunt, C., Moore, K., Hurst, S. M., Lucas, R., Roche, M., Gaillardin, C., Tallada, V. A., Garzon, A., Thode, G., Daga, R. R., Cruzado, L., Jimenez, J., Sanchez, M., del Rey, F., Benito, J., Dominguez, A., Revuelta, J. L., Moreno, S., Armstrong, J., Forsburg, S. L., Cerutti, L., Lowe, T., McComb, W. R., Paulsen, I., Potashkin, J., Shpakovski, G. V., Uesery, D., Barrell, B. G. and Nurse, P.

The genome sequence of *Schizosaccharomyces pombe*
Nature 415 (6874), 871-880 (2002)

21848401

11859360

2 (bases 1 to 36315)

McDougal, R. C., Rajandream, M. A., Barrell, B. G., Brown, S., Murphy, L., Jones, L., McNeil, A. and Harris, D.

Direct Submission

Submitted (17-JAN-2000)

European Schizosaccharomyces genome

sequencing project, Sanger Institute, The Wellcome Trust Genome

Campus, Hinxton, Cambridge CB10 1SA, E-mail: pombe@sanger.ac.uk

Notes:

Details of *S. pombe* sequencing at the Sanger Institute are

available on the World Wide Web.

(URL, <http://www.genedb.org/genedb/pombe/index.jsp>) (URL,

http://www.sanger.ac.uk/projects/s_pombe/)

CDS are numbered using the following system eg SPAC5H10.01c, SP (S.

pombe), A (chromosome 1), c5H10 (cosmid name), .01 (first CDS), C

(complementary strand). However, clones may have been reorientated

since the original submission, therefore the complementary strand

notation may be invalid for strand inference. IMPORTANT: This

sequence MAY NOT be the entire insert of the sequenced clone. It

may be shorter because we only sequence overlapping sections once.

Location/Qualifiers

1..36315

/organism="Schizosaccharomyces pombe"

/mol_type="genomic DNA"

/strain="972h"

/db_xref="taxon:4896"

/chromosome="1"

/map="IR"

/clones="cosmid c30"

/complement(253..5721)

/gene="SPAC30.01c"

/note="synonyms: sec702, sec7b"

/complement(253..5721)

/gene="SPAC30.01c"

/note="guanine nucleotide exchange factor (predicted);

involved in the secretory pathway (predicted); Sec7

domain; similar to *S. cerevisiae* SEC7; non-clathrin

vesicle coat component (putative); involved in

intracellular protein transport"

/codon_start=1

/protein_id="CAB66460.1"

/db_xref="GI:6723884"

/translation="MQDASPEVIMNPAIPWNSQEHDRHEDPEVCTTSVTKIINDS

DNEDDNGMDLGRVASESLEDGAVVSDINTESSLSIPAQENKSEGEIQKQED

LKDKNQETIATVPDPAASALEKQGLNLEKSTHVLLIKIQLQIKVCMQN

EMKMLQKPLQTCQDFKFLVLFQYCFSSQDEVMNLSLDTISKLAPAYS

SKDTPASFGPKSLIQCMQVDCSDINDEVDGNLQVNVKALSAFILCSQDSMLH

GAILNSVRKLFNVPLIGDSDTIQSVAQSLTQAVTVVERLRASHTSQNSTALPEE

DASVTENWHDDEDPDKITLHSMASAGTSLDHVKVDADDPAVTSVENSIIQDAFLV

FRSCLRAVQTSQKVSINRSQMRALISLHLIYRILEKNSDLFMDPTIQFGIPA

LKGMILVHARQYICLVLSRNASVPQVFEVCCDIFYLVFSLRAHFQKEIVFFPE

VYFPMPLDKNTSNOKLTLIIQICLNPRALVELIYNDQSRSTNTVFQQLLSI

SKVTNGPSETISDEIILPSLESRSSTPFLNTNSASLSEVQLTITFSFQQLKL

KTLCVLDILQSLNWAESGLYSRRGVSTDEQGFVGDYDALSRSDPTVPTNYNGQ

SFEANSHSIIALADPSQESNQRKLLARTCLINKNYKPTKGLKMLSENYVDIND

PKALAEFLFRADGDKTTLGLDYLGEQDEKISVMVHEFDICLSFINLAFVADLRRLQD

FLPGEAQKIDRLMKFSERYMKNPSAFANADTAYILAYSIIILNTDLSPRIKKN

TKEDFIKNRGINDDADLDEYLGIVYDDILKNEIAMKDDQELAAIAPLMNFTSSG

FTFTSNRDLQRVACIQASEMANKATSVLKLLYQOKHGSQKTNVYNATHEFHIG
PMLEATWNPILAALSPLONSDYVNELNMLCGLDQVLRVIACLPDLDIRDAFKTLT
NFTNLHSTSEIKLNTWNTKLLTARIAGEGNLADSKMDIITIIISQERVLQIGVUD
ETEVPDYNARVRKNVNGSSINSIRHVSGSTSRSTRSLKSLSEANVELMSTEV
VLSIDRITQTSLSGSAIVFFKALCEVSWDEITSSDLEQPRLYISLQKLVESLYN
MQRIRWSSITWNVLGFRFNNWSDENRHVAFALDSLRQSMHFLTEIELSLFSQK
EFLKPFYVNASDTPVEVKELQVQKMQIKSKI KSGWKTLFGVPTFAKARSEI
LISMTFTVLNLFSEHDTLMQCNLDMLISFTLCKNGTNQKISLQSLRIKREVS
SLSTMIKEGSSKSVNETSKYVPVLFAYVDIIMSAEDLEVRSLQNLVLEALFL
SDDTESTVEVSKRFIFPISIFGPEADEATVMLRBEERTWSTLVEALRLVLTL
LTRFRDKHLLKGLVFNCSICRDNITLSRGTNCNQQLSGNARFVKDNLVA
DMFLFKETTPHQLLLETSFGQGPVSENETQLSHRGSGSLPETSISTSI
SPEKQFRSMIRKCIILQLLISIVABLLNEEVFNHPHEHLKITVAVDSQFAR
KFNEKSLRITLLNVGFMKPNLLROETASALLYITLPRLLKTRDPLGKTETDOKI
HKLKFPVCAEMLDYASLVKVEKTRNHAQPVATILDSILNLPLELSENIHTLYF
SCCSMIAKENLDQRLKKNYFNRVGHILLNNAQGE"

complement(3049..3609)
/gene="SPAC30.01c"
/note="Match to PF01369 Sec7, Sec7 domain Score 329.12"
complement(6110..6195)
/note="mRNA from AU011189"
complement(join(6196..6869,7015..7186))
/gene="SPAC30.02c"
complement(join(6196..6869,7015..7186))
/gene="SPAC30.02c"
/EC number="2.7.1.78"
/note="polynucleotide kinase (predicted); RNA polymerase
II elongator associated protein (predicted); similar to *S.*
cerevisiae Kri12"
/codon_start=1
/protein_id="CAB66461.1"
/db_xref="GI:6723885"
/db_xref="UniProt/TREMBL:Q9P7V4"
/translation="MPLIIVSGYPSSTKTRSNELKALEDRHONIDNTKDYRVIII
NDESLNEKETRESKNEKARGLLYSVAVERLSKSTFVICDALNYIKGRFYQYCES
KSMYTHCVIHVAPQDLCKRKNSEQPYDDVLEQMFREPEGNTRWDSPLEFTV
LHDDACPIDDIWSVLIHNKVPKNQATVVKPAEVNLYELDKTQDQVIMLIDNSN
DTSILITVPGSKLQIALPSVTVSLPQLRLRRQFIQNRQSYNTNVLKEMPVEFLNGQ
FETLD"

complement(6870..6883)
/gene="SPAC30.02c"
/note="ctaattgtcttag, splice branch and acceptor"
complement(7009..7014)
/gene="SPAC30.02c"
/note="gtatgt, splice donor sequence"
complement(join(8004..8038,8096..8264,8307..8389,
8428..8788,8827..8889))
/gene="SPAC30.03c"
complement(join(8004..8038,8096..8264,8307..8389,
8428..8788,8827..8889))
/gene="SPAC30.03c"
/note="translin-like protein; DNA-binding protein;
involved in meiotic recombination (predicted);
non-essential (PMID 12618370); no apparent *S. cerevisiae*
ortholog"
/codon_start=1
/protein_id="CAB66462.1"
/db_xref="GI:6723886"
/db_xref="UniProt/TREMBL:Q9P7V3"
/translation="MKNSTFIQDQDQDKHSIREKLTAEVLDLDEKRLVQLLLANC
EQSRNLEQHEGLTLEDEQNEILEALEIKSKTRGLAELASFPYKYNGWDR
SIVKVVYLLASWTGRDKSLRPTYSLSLSEVGQILOVPFPEESTFHLISBOYLH
AQLSCSELARQSVNSVIGSNYHIFPEALNTIKVHSSFQVLSKNDLSRRHPDLKY
DLKRSDEVYDLRIHKLIV"

complement(8039..8051)
/gene="SPAC30.03c"
/note="ctgacatagcag, splice branch and acceptor"
complement(8090..8095)
/gene="SPAC30.03c"
/note="gtatga, splice donor sequence"
complement(8265..8280)
/gene="SPAC30.03c"
/note="tgacacaattcttag, splice branch and acceptor"

misc_feature

3'UTR

gene

CDS

misc_feature

misc_feature

gene

CDS

misc_feature

misc_feature

misc_feature

```

misc_feature complement (8301..8306)
/gene="SPAC30.03c"
/notes="gratata, splice donor sequence"
misc_feature complement (8390..8403)
/gene="SPAC30.03c"
/notes="ctcacaattctcag, splice branch and acceptor"
misc_feature complement (8422..8427)
/gene="SPAC30.03c"
/notes="gtgcgt, splice donor sequence"
misc_feature complement (8789..8800)
/gene="SPAC30.03c"
/notes="ctaacattacag, splice branch and acceptor"
misc_feature complement (8821..8926)
/gene="SPAC30.03c"
/notes="gracta, splice donor sequence"
gene complement (9229..13638)
/gene="SPAC30.04c"
CDS complement (9229..13638)
/gene="SPAC30.04c"
/notes="ABC transporter family; unknown specificity; 15
predicted transmembrane helices; similar to S. pombe
SPAC989.12c"
/codon_start=1
/protein_id="CAB66463.1"
/db_xref="GI:6723887"
/db_xref="GOA:Q9P7V2"
/db_xref="UniProt/TrEMBL:Q9P7V2"
/translation="MENFHRPFGGFGVRVTSYLSLDFSLSAISIFPHYDQP

Alignment Scores:
Pred. No.: 0.688 Length: 36315
Score: 179.50 Matches: 131
Percent Similarity: 37.34% Conservative: 124
Best Local Similarity: 19.18% Mismatches: 289
Query Match: 5.45% Indels: 139
DB: 8 Gaps: 28

US-10-736-868-2 (1-643) x SPAC30 (1-36315)
Qy 17 ProLeuSerAlaGlnSerProSerThrSerAspAlaProGlyAlaLeuLeuSerSerLeu 36
Db 34212 CCTTTTCAGCGCTCTTATCCCTCTAAAGCAGATGAGAAAATAGCTATCATCTTGATCTT 34153
Qy 37 ValGlyLysSerHisGlnLysLeuPro-----LeuAlaProSerMetGluAla 52
Db 34152 TATAACTTCGGGAGTAGTTAAACGGTGAAAGGAGGTATCTTAACGGTAGCGGATGACCTA 34093
Qy 53 LeuGluLeuMetGlyValGlnPheValAspAlaLeuLeuLysGlyGlnMetGluMet 72
Db 34092 TTGAAGAAATGATGGGAAGAAGTTATAGAATGATGGAACAGTTAGCTGAGCGCTGTAIG 34033
Qy 73 AlaLys-----GlyAlaPheLysThrGlnLeuGluVal 83
Db 34032 CAGCGTGAAGATAATAGCAATTTTACGAAACGAGATTATACGAAAGCGGTTGGAATAT 33973
Qy 84 LeuGluLysValHisProAspGlnPheAspLysTyrLysLysLeuLysValAsp----- 101
Db 33972 GACGAAGATGAAGAAGACGAGAGAGAT-----GTAGATGAGGAC 33931
Qy 102 -----AspLeuAlaAlaAspAlaValMetGlnGlnAlaGluMetAlaLysLeuGlnPro 119
Db 33930 GAGTTAGATCTAATGACCGATGAACAAAGGATGGAAGAGGACGAGGATGTTCCAAATT 33871
Qy 120 LysSerGlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIleProIleGlySerSer 139
Db 33870 TTTGCTCCCGTTTATTGAGCAACGTTG----- 33841
Qy 140 IleArgGlyLeuGluAspAlaIleArgThrGlnArg----- 151
Db 33840 TTGCAAGCTTATCGTGAGAAGTTGCTCAACACGCGCAAGCTAACTTTTAGAAGAGATT 33781
Qy 152 AspMetGluAsnThrAspProSerGluGlnIleAlaLysAlaValMetAspLysPheGln 171
Db 171

```

```

Db 33780 GAGGAGGAGAACAAACGCAACAGGAACGCTCAGAGCTCAAAAAAATAAGAGAGAGAGAAA 33721
Qy 172 ThrGlnIleLeuProGlyLeuValAlaAsnMetIleAlaGlyLysAsnProPheLysMet 191
Db 33720 AAGAGG-----GATAAAAGAGAGCAATTGAAATTG 33691
Qy 192 ProGlnGlnMetArgLysAlaGlnAlaAlaProSerSerValPheGlnGlnAlaLeuAla 211
Db 33690 GCGAAGGAGAGAGCGTCAAGAGCGGAAGCTGAGCGTCTTCTGCAACAGGCT---GCT 33634
Qy 212 GlnArgAlaMetLeuGlyLysAsnAlaProValAlaGlyGlyArgGlyGluGlnArg 231
Db 33633 CAAAAGCATTGGAAGCTTAAAGACAGGAAGAGCTCGCAAGAAACGCGGAAGACCAAGT 33574
Qy 232 MetMetMetAsnArgValAspGlnArgMetGlnGln-----ArgGluLeuGlnGlu 248
Db 33573 -----TTGAAGCGTGAGCAGGAGAAAAAGCAGCAGCAAGACTTGAGCGTCAGAAAAAGAGAG 33520
Qy 249 GluAspGluAspAspAspLeuGluAspGluAspValProArgArgArgSerSerAsp 268
Db 33519 GAAACAAAGCAGAGAGGAAGAGAGAACTGAAGAGCAGCAACAGAGGTGAC 33460
Qy 269 GlyGluProGlnSerGluAlaGluHisGlnArgAspLeuAlaArgLeuLysSer 288
Db 33459 CGTGAATAAATGCGGAGAGACAACGACTAAGAGAGAGAGAAACGTTATTTGGAA 33400
Qy 289 SerProArgLeuLysGluLeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrGln 308
Db 33399 GAAAGAGAGGAGGAGGAAAAATTAGACAAGAGAGAGAGAA----- 33358
Qy 309 ArgMetArgAspSerProLeuSerLysArgArgProGluAlaMetAsnAspGluAspGlu 328
Db 33357 -----CGTGGCGTCTGAGTTACTTCTCAACTTGGCGCTTTTGAAAAAAGCGCAT 33199
Qy 329 SerAlaPheArgAlaMetGluAlaArg---AlaLysLeuAspGlnLysSerGlnLeuVal 347
Db 33318 AAAAGAAAGGAGATTAAAGAGAGCTAAGATCGCTCTTCTTGCACCTAATCAGACTAA 33259
Qy 348 LeuGlyLeuHisGlyPhe-----GlyGlu 355
Db 33258 GAGGGTTCGATGGTTGTACTTCTTCTCAACTTGGCGCTTTTGAAAAAAGCGCAT 33199
Qy 356 SerAspAspAspGluAspGluAspGluAspGluAspGluAspGluAspGluAspGlu 375
Db 33198 CTCGTAACGATGAGATGAAGTTCTTTCACATCTCTAGATTCTGTACTAATGCTTCT 33139
Qy 376 ArgArgAlaProLeuArgLeu-----SerSerGlyPheValGluLysLysLysSerAsn 393
Db 33138 CGGCAAGCTCCCATAGGCGCTTAAGAACACAAATAATTATCGGAGCGCAATGCTTCTTCT 33079
Qy 394 AspGluLeuLysSerAlaLeuAspArgIleLysTyrArgValAspAspValGluLysTyr 413
Db 33078 AATCTCTTAATTCTTCGTTATTTTCATCTTTTAATTCGCTCAACCTTTAATTCT--- 33022
Qy 414 LeuAlaProLysProMet-----GluPheAsn 422
Db 33021 CTGACCTTAATCTTAAATAGACACTTAATAATTCGTAACTAATCTGATTTGGT 32962
Qy 423 ProLysProGlnProGlyTyrPheAlaProArgLysIleProThrArgProArgLysMet 442
Db 32961 AGAAGGCCAAAT---GGTTTACATCTCTCATCTCTCACTCTCTAAGTAATTCACCAAT--- 32908
Qy 443 LeuProLeuLeuIleGlySerAspProLysValGlnGluGluIleArgArg-----His 460
Db 32907 -----TTTGGATTAAATCCAAATGCTCGGCATTTCTTTAGTAGGCCAATTC 32860
Qy 461 ProSerThrGluTriphysIleAla-----LysGluSerArgValLeu 474
Db 32859 CCGGTGCATCATTTATCTTTGCAACACCTCTTCTCAACGTCGCTAATAATATCCATTG 32800
Qy 475 ThrAsnLeuLysAsnAsnProSerLeuAlaLeuPheMetAspAspLysLeuGluAsn 494
Db 32799 AATAATGGAGCAACGATACGCTTGTCTTAATTCATTTTCAAGTCCCAATTAAGTCCT 32740

```


Db	1477	ACAAATCAGTCCTATTATTTTCAACAACAACAACAATCCATCTCTTTTGTAGTACAGTGGTT	1536
Qy	489	AspAspIysLeuGluAenThrLeuLysGlyArgGlnMetLeuThrAspGluGlnLysGly	508
Db	1537	---GGAAGGTTAAACAACATTAGGTGGGAAAGAGCCGCAAGAAAGATTACACAAGACT	1593
Qy	509	-----ArgThrArgValIysThrIleArgAlaLeuProArgLeuPhe	522
Db	1594	ATGGAGTCTATAATCTTTAAATAACAAGGGTAAAACTCTT-----	1632
Qy	523	GlyAlaProThrAlaLys-----AlaGluMetIleAspAlaLysValPheGln	538
Db	1633	---CTAGAAACAACAAGAACCTTCAATGCTCTGAACTGGTGAAGTT---GTGTTCCAA	1686
Qy	539	AspIleGluGluArgProIleProPro 547	
Db	1687	GAT---CCAGAAACCAACGAGTAAACCA	1710
RESULT 11			
LOCUS	AF369029/c	292967 bp	DNA circular VRL 26-JUL-2001
DEFINITION	White spot syndrome virus, complete genome.		
ACCESSION	AF369029		
VERSION	AF369029.1	GI:15021392	
KEYWORDS			
SOURCE	Shrimp white spot syndrome virus		
ORGANISM	Viruses; dsDNA viruses, no RNA stage; Nimaviridae; Whispovirus.		
REFERENCE	1 (bases 1 to 292967)		
AUTHORS	van Hulten,M.C., Witteveldt,J., Peters,S., Kloosterboer,N., Tarchini,R., Fiers,M., Sandbrink,H., Lankhorst,R.K. and Viak,J.M.		
TITLE	The white spot syndrome virus DNA genome sequence		
JOURNAL	Virology 286 (1), 7-22 (2001)		
MEDLINE	21342572		
PUBMED	11448154		
REFERENCE	2 (bases 1 to 292967)		
AUTHORS	van Hulten,M.C.W., Witteveldt,J., Peters,S., Kloosterboer,N., Tarchini,R., Fiers,M., Sandbrink,H., Lankhorst,R.K. and Viak,J.M.		
TITLE	Direct Submission		
JOURNAL	Submitted (15-MAR-2001) Laboratory of Virology, Wageningen University, Binnenhaven 11, Wageningen 6709 PD, The Netherlands		
FEATURES	Location/Qualifiers		
source	1..292967		
	/organism="Shrimp white spot syndrome virus"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:92652"		
gene	1..615		
	/gene="VP28"		
	/note="envelope protein"		
	/evidence=experimental		
CDS	1..615		
	/gene="VP28"		
	/note="envelope protein"		
	/codon_start=1		
	/product="ORF1, VP28, gene family 1"		
	/protein_id="AAK77670.1"		
	/db_xref="GI:15021393"		
	/translation="MDLSFTLSVVSAILAITAVIAVIFIVFIRYHNTVTKTITHTDNI ETWNELRLIPVTAEVGSGYFKMTDVSFSDTLGKIKIRNGKSDAQMKEDADLVITP VEGRALVTVGQNLTFEGTFKVNNTSRKINITGMQVPKINPKSKAFVGSNTSSFTP VSIDEBVGTFCVCGTTFGAPIAATAGNLFDMVHYVYSGTETE"		
	complement(710)..2902)		
	/codon_start=1		
	/product="ORF2, putative serine/threonine protein kinase (PK1), gene family 2"		
	/protein_id="AAK77671.1"		
	/db_xref="GI:15021394"		
	/translation="MEGGDQRTKLTPATVMGLVQSKTPGEGEGGGQFKIPSAIAY KSCCKNATRRSPSDSPYSRLPMKLNKNNGEVGGKAPPPVTLRLURFDEYESTPINFN RNKKRPITIDENQAFATNPTATDIKKQQLPVSVAASVLRKHRANADTQYRRFSH PNCAKFTDNLKARDYTPLSVLRSHVGPKLKSSCDTVTETNVKRNFSIDKWKVL EKPCPYFAVEADTNIAGLESPPHILROAKGLGISDVODVSNTYETIKQSCIDAKE KASKFLWSNNRTKQPPSSWVPVFGSKNLVLDTSPLLNWRNLCKNGKGWIKMTMSID		

HMAKNVFKLSPGACESILLEKTTLLGEVTAQCKKWEYRRNIPVPAHVQVPOVASOVVM			
IGPSELYLEVKGVYVMLETKVIFKMTDKENYCEVFETVESHALBEGMKAGVGRK			
MCVEGFQEMDFAGISVDVNLGDLCKMDENVVOQNPSTTSSRPAELMQLDQHSGLC			
NMRDLYGVYMLQATGRUPEGLQSKKPKPITDSISAIAIVGKMRMLNQLPFLVVEI			
NVTRLSSQGLVNPDIKSDNTVIDGITQPRMIDFLIVPCKYKYNFCWGTDERFF			
SNHPTAPFINSLSCTSETMTFGLAYLLIDMLSLIKRTADLSANSITVNTIPLSLIV			
SNWYDQEKTRPRAYEIAPIVIGACFPFKDNIAKLFQSPKHSLSYKVKV"			
complement(3118)..4989)			
/codon_start=1			
/product="ORF3"			
/protein_id="AAK77672.1"			
/db_xref="GI:15021395"			
/translation="MAWTVMALKDAFTERLVVNVKSGTDMAPVVEDDDQKSLFOKVE			
NLYRVLVVEKNSALTLSGNKNTKROQVEEDKVIPEGEDRTVSNLPOAKETIAA			
NAESILDYWKVNPFLDDTKKRSKSDTFLRTAVICLVRCVSKYDKMTCSLISYPE			
HKILNKSLDPLDILDNKQELLHMDSKYGKTTSPELAKETIEALYTTIVVHNWNA			
FKLYQASLTHKPTVKKVASVIFRTWRKIYKAVYKNNVERDLNKNIMKNEAD			
NANVLTIKMYKKIGNSVNTNNNSAHQMSDESDDDDDCEGDMDCEASEREKK			
HOESLYPINTVTTIGDYIFKVLLELVSPHHPWKIPMCDFVNRTNPKLMKAMET			
DISNAIVEASKNVPQIILPIAANFMDFCKSGPSPDYKFCMNEPSSNETLSSGA			
GVYGFPGFPFSGHKKELDIIISNLSRLNLNKEADNLSTRIWREGSVVCFNCPITA			
RGAVLGYQLSRSIKALWAKKIODAVTESVKQRQNAADKNSRNCDDLGEDEVSMK			
TVTFGCANMLKTQNGMGKFNVVVSPEDSITQANKEGAARQYMSQQVFTHSPALDQSK			
5185..8970			
/codon_start=1			
/product="ORF4"			
/protein_id="AAK77673.1"			
/db_xref="GI:15021396"			
/translation="MASSGGFTGIDDLFKTVIOBEKQKKNKPTQAPETEPKPGPSQA			
PPDPVPKPTPTNFCPPPPNPPLPPPPPPKPSREERLKTSLRINKALSIDIVEATN			
BRVDALKENQALNTEYDKNDYFVLKCSIITPSVPTAIIGAHVQVAKSEIBLANVE			
LDIKKCSLYVNNESLKFPRHENLILQIAVQLFSRHNTKCVGAEICVGNKKNKF			
VNKLNVKKNAPSSSTVLEIRGATRNLENNFNNGENNTVNNKDIIPSERANLDT			
TKAEISHVFSTLHRDTSKDLFFKNTFYORPFTDNKFRWTEVIQWTESEASKQITK			
SLDKPTDNLFLVPHSNNLADHLRKFKNVLYKNSHAKRNYTKQETELNPQID			
SAKEYKMFAEIDKCLDVLLAIGKNDKYTKSTVIOYGRKFRYLIICYAFYALNKAH			
SRVPSLPENFELISFMYCHGPFHLSASFLSTLTVEYOHMPFMGTAAPSVSAKLM			
DIDSALMKGGKGVGRDFGSPKTSLSHTTLYSLFGLFAEMAGMTMTALLSGHNVKYP			
ALQORISLERWCDSVIFITFLYFHRFGSAKKVLSLESALPLINGQTHANTKRYRA			
AKRCIEAAMEGVEEAGLTLSYAHLLGPSIQKALGLPVLKINLPKINLTAASSQTN			
LGDFFVEQLLAKKAFEPFAEGTAGFLGMDNLVKDSIDKYTGEGGAFSDVENVKQM			
EONTPYDTSALMTPIPKAFYBEEDVPQOENSTQORYLNRDVEEYLMASPMKVF			
VSILDKTNOKERFMSVGDIALLAWCNRVNLKKNWEYAIKNGYEWLGAKMCHLLIL			
ADLVNFGILGLKINTKLDNTDTHRSDRLPSVADQKFKINTLSLDRKQALAVHS			
CNVSTRTHVGTATSWAIDALRTYTRGDKMFAALSSLDMLHGLHNTSANEVYF			
SNLYLCNEQENGLWGYTRTSEKJAKBELGRGLGNKVGAKTLELAALAAIAISAL			
DMGEVANEVSDSSVKRIASTCLNNAKVAASAREKAREASIKRLLLATNAPAGSSR			
NSNRFLIKDLWGFSDPDQRKQLIKGEAVSLCPNTGFLHAAVDPDIYFYSFSETSI			
VLRLRLIKPEQIDEMVCPSTAPEANKRKLVRNNQDAVLTLDDENIVKYNKYDMVE			
DEARERLRHQDKQSVIAARISKVCERKNPKKRRLELQSDVDEQLIRELAAIAY"			
9056..10879			
/codon_start=1			
/product="ORF5"			
/protein_id="AAK77674.1"			
/db_xref="GI:15021397"			
/translation="MEDDTFNOETAVKLVRYWTEYDCCCLPLVNRMBRLILGSGGVDA			
TSVRSPALYEEDKKGDKCIPRITSLSIGILLERALTDPDLAAAFADYSEKLVYCS			
NNTQGNFDRYSMTIWLIDGNNSKRYEVTCPCTVEKISGGAESLHKPMSLLAIFNNIL			
KEAFAERIELKKLILSLTSGAAGGCMYKDSQSSQSGNSWTSLFHTSKDKTLRL			
YVILVNSKIKHTSRQPCVSDLLYALCSTNNASAYAKARNLCVLEGGFLVLPK			
YTIIEBNGPFDKTDIQLSVNNEPVSETNSSLAAASSSLEDCCDDDDDDDDDD			
EKTKKQPKKQTKKQTKTTSTLPLPISKTNDNMLNVLKKA VNGKRKMWLSLGGKG			
QHSKKLTSAAAGGASSDVVAGENEENPSSVSTNNRDKDYLPCLPQIIEVITF			
SOHRMNNKLAEVSHVSHVINGENCLNLTQHRKYLILPHENLILCPFLQVHVGFK			
FRILTGVSCFFDRIEIVFSDQSDSVLNNAAHSAILRLLSYIRENSLRKSVRTASVK			
GIDFVVKQSQDTNIGIPLSNKEIRERQLCSASTLSMLGLGK"			
10834..13236			
/codon_start=1			
/product="ORF6"			
/protein_id="AAK77675.1"			
/db_xref="GI:15021398"			
/translation="MLSPNPEYAFWFGKMTDPGVILPVSKOVYLFSGRSQSDVGINT			
LDPHDLIDIKITSKRIGVERLAQYNLPDMDFRAMEKLNLSRNMKESIPTGIFLDTG			


```

QY 290 oArgLeuLysGluLeu----- 295
Db 4987 GGAACCTGAAGGACTGGAGCGCACATCGACTCGGCCAACAGAACCGGGACGAAGCCAT 5046
QY 296 -----LeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrGlnArgMetArgAspse 313
Db 5047 CAACAGCTCGGAGAGTGGAGCGCCAGATGAAGGACTGATGGCGAGCTGGATGACAC 5106
QY 313 rProLeuSerLysArgArgProLeuAlaMetAsnAspGluAspGluSerAlaPheArgAl 333
Db 5107 CCGCGCCTCTCGTGAGAGATCCTGGCCACAGGCGCAAGAGACGAGAGAGTGAAGAG 5166
QY 333 aMetGluAla-----ArgAlaLys-- 339
Db 5167 CATGGAGCCGAGATGATCAGTTGACAGGAGAACTGGACGCCGCGGAGCGTGCACAGCG 5226
QY 339 ----- 339
Db 5227 CCAGGCCACGAGCGGGATGAGCTGGCTGACGAGATCGCCAAACAGCAGCGCAAGG 5286
QY 340 -----LeuAspGlnLysSerGlnLeuValLeuGlyLeuHisGlyPheGlyGluse 356
Db 5287 AGCCTGGCGTTAGAGAGAAGCGCGCTGAGAGCGCCGATCGCCAGCTGGAGGAGGA 5346
QY 356 rAspAspAspGluAspGluAspGluAsnLeuLeuAspProSerGluAsnSerPheAr 376
Db 5347 GCTGGAGGAGGAGCGGGCAACAGCGAGCTGATCAACAGCCGGCTGAAGAAGGCCAACCT 5406
QY 376 gArg-----AlaProLeuArgLeuLeuSerSerGlyPheValGluLysLeuLys 391
Db 5407 GCAGATCCACAGATCAACGCCGACCTGAACCTGGAGCGGGCAGCCGACAGAG----- 5461
QY 391 sSerAspAspGluLeuLysSerAlaLeuAspArg-----IleLysTyrArgVa 407
Db 5462 ----AACGAGAATCTCGCGCAGCAGCTGGAAACGCCAGCAACAGGAGCTTAAGGTCAAGCT 5517
QY 407 lAspAspValGlu-----LysTyrLeuAla- ProLysPro----- 418
Db 5518 GCAGGAGATGGAGGGCACTGTCAAGTCAAGTACAGGCTCCATCAACCCGCTCGAGGC 5577
QY 419 -----MetGluPheAsnProLysPro-----GlnProG 428
Db 5578 CAAGATTGCACAGCTGGAGGAGCAGCTGGACACAGGAGCCAGGAGCGCCAGCAGCCTG 5637
QY 428 lyTyr----- 429
Db 5638 CAACAGGTGCTGGACCGAGGAGAGAGCTCAAGATGTGCTGTGCTGAGGTGATGACGA 5697
QY 430 -----PheAlaProArgLysIleProThrArgProArgLysMetLeuProLeuLys 448
Db 5698 GCGAGGAAACGCGGAGCAGGTACAGGACCGAGCGGCGCAAGAGCATCTACCCGCTGAAGCA 5757
QY 448 lySerAspProLysValGlnGluLysIleArgArgHisProSer----- 462
Db 5758 GCTCAAGCGGAGCTGGAGGAGCGCG--AAGAGGAGCGCCAGCGGGCCACGCGCTCCCG 5814
QY 462 ----- 462
Db 5815 CCGGAAACTGCAGCGGAGCTGGAGGAGCGCCTGAGAGCGCGCATGCCATGAACCGCGA 5874
QY 463 -----ThrG 464
Db 5875 AGTCAGCTCCCTAAAGAACAAAGCTCAGCGCGGGGACCTGCGGTTTGTGTCGCCGCCCG 5934
QY 464 luTrpLys-----IleAlaLysGluSerA 472
Db 5935 AATGCCCGGAAAGCGCGCGGGATGGCTCCGACGAGAGGATAGATGCGCAAGCGGATGG 5994
QY 472 rgValLeuThrAsnLeuLysIleAsnAsnProSerLeuAlaLeuPheMetAspAspLys 492
Db 5995 GGCTGAGGCCAAACCTGCCGAATAAGGCTCTTCTCTGACGCGCTGAGATGAGGACAGA 6054
QY 492 euGluAsnThrLeuLysGlyArgGlnMetLeuThrAspGluGlnLysGlyArgThrArgV 512

```

```

Db 6055 CAGACACCAAGCCTCCCTCCAG-----ACCCGCGAG----- 6089
QY 512 alLysThrIleArgAlaLeuProArgLeuPheGlyAla-----ProThrA 527
Db 6090 -----CAGGCTCTCCCACTTCTTGGGACTGCTGGAACATGCTCTCTCTG 6138
QY 527 lalysAlaGluMetIleAspAlaLysValPheGlnAspIleGluLysArgProIleProp 547
Db 6139 CCTC-----CGCCCGCTCCCTCC-----CGCCCGCTCCCTCC 6156
QY 547 roLeuPhePheGluProLysGlyArgHisThrArgLeuArgTrpThrGlyAlaAsnGluL 567
Db 6157 CATCCCGTTCCCT----- 6170
QY 567 yseLulleProGlyLeuGlySerArgPheIleLeuProSerLeuAspPro-----T 584
Db 6171 -----CCAGGTGTGTGAGGCACTTGGCTTCTCTGCTGCATCCCTTCCAGCTC 6222
QY 584 hrMetProAlaLeuAsnThrAlaPheSerThrGlnGlyArgAlaArgAspGluTrpAspT 604
Db 6223 CTCCCTCTCTCAGATCTGATACCAAGAGACAGCGCGCGGCGCAGGAGAGCGACC 6282
QY 604 hrMetPheLysIleProAsn 610
Db 6283 AGCAGGCTCTCAGCCCTCT 6302

RESULT 15
LOCUS HUMMYONM 5122 bp mRNA linear PRI 27-APR-1993
DEFINITION Human nonmuscle myosin heavy chain (NMHC) mRNA, 3' end.
ACCESSION M31013.1 GI:189035
VERSION M31013.1
KEYWORDS myosin.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5122)
AUTHORS Saez,C.G., Myers,J.C., Shows,T.B. and Leinwand,L.A.
TITLE Human nonmuscle myosin heavy chain mRNA: generation of diversity
through alternative polyadenylation
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 87 (3), 1164-1168 (1990)
MEDLINE 90138958
PubMed 1967836
COMMENT Original source text: Human macrophage U937, cDNA to mRNA.
Draft entry and computer-readable sequence for [1] kindly submitted
by L.A.Leinwand, 21-DEC-1989.

FEATURES
source
    1..5122
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    <1..5122
    /product="NMHC mRNA"
    <2..3745
    /notes="nonmuscle myosin heavy chain (NMHC)"
    /codon_start=1
    /protein_id="AAA36349.1"
    /db_xref="GI:189036"
    /translation="QSFRRORYEILTPNSIPKGFMDGKQACVLMIKALELDSNLYRIGQ
    SKVFPAGVLAHLEERDKITDVIIGFOACCRGYLARKAFKAKQKQDAMTKVLOQRNC
    AYLKLRNNQWRLLFTKVPKLQVSRQSEEMAKKEELVKVREKQLAAENRLMEMETL
    HQAEKKKQVQNIQEELQEEESARQKLEKVTTEAKLKKLEBEQIILEDONCKL
    AKELLEDRIAEFTTNLTTEEBKSRLAKLNKHEAMITDLERLRERKEQKQLEIK
    TRKLEGDSTDQIAELQAOIAELKLMOLAKKEELQALARVEEAAQKNMALKKI
    RELESQISELDLESERARNKAEKQKDLGLEELALELTDLDSTAAQQLRSK
    RQEVNLIKLTLEBEAKTHEAQIQENRQKHSQAVELAEQLEQTKVKANLEAKOTL
    ENERGLANVEVLLQGGKDSHRKKRVEAQQLQVKKFNEGRVETELADKVTKLQV
    ELDNVTGLSSQSDSKSLTKQFSALESQLODTQELLQENRQKSTLTKLKQVEDEK
    NSFREQLEBEAEKHNLEKQATLHAQVADMKKMEKDSVGCLEAEVKKRKLQKOLE
    GLSORHEEKVAAVDKLEKTKRLQQLDLDLVDLHQRQSAACNLEKKQKFKDQLLAAE

```



```
Db      3478 GAAGCAGCTCAAGCGGAGCTGGAGGAGCGCG---AAGAGAGGCCCGCGGGCCCAACGC 3534
      |||      :|||      :|||      :|||      :|||
Qy      462 ----- 462
Db      3535 CTCCGCGGAAACTCGAGCGGAGCTGGAGGAGCGGCACTGAGACGCGCGATGCCATGAA 3594
Qy      462 ----- 462
Db      3595 CCGGAAGTCAGCTCCCTAAAGAACAGCTCAGCGCGGGGACCTGCGGTTTGTGCTGCC 3654
Qy      463 --ThrGluTrpLys-----IleAlaLysG 470
      |||||      :|||      :|||      :|||
Db      3655 CCGCGGATGCGCGGAAAGCGCGGGGATGGCTCGACGAAGAGGTAGTAGCAAGC 3714
Qy      470 luSerArgValLeuThrAsnLeuLysAsnProSerLeuAlaAlaLeuPheMetAspA 490
      :|||      :|||      :|||      :|||
Db      3715 GGATGGGCTGAGGCCAAACCTGCCGAATAAGCCTCTTCTCTGCAGCCTGAGATGATG 3774
Qy      490 spLysLeuGluAsnThrLeuLysGlyArgGlnMetLeuThrAspGluGlnLysGlyArgT 510
      :|||      :|||      :|||      :|||
Db      3775 GACAGACAGACACACAGCCTCCCTTCCAG-----ACCCGCGAG----- 3815
Qy      510 hrArgValLysThrIleArgAlaLeuProArgLeuPheGlyAla-----P 525
      |||      |||      |||      |||
Db      3816 -----CACGCTCTCCCGCTTCTTGGGACTGTGTGAACATGCTC 3858
Qy      525 roThrAlaLysAlaGluMetIleAspAlaLysValPheGlnAspIleGluGluArgProI 545
      |||      |||      |||      |||
Db      3859 CTCTGCCCTC-----CGCCCG 3876
Qy      545 leProProLeuPhePheGluProLysGlyArgHisThrArgLeuArgTrpThrGlyAla 565
      :|||      |||
Db      3877 TCCCCCATCCCGTTCCCT----- 3896
Qy      565 snGluLysGluIleProGlyLeuGlySerArgPheIleLeuProSerLeuAspPro--- 583
      |||      :|||      :|||      :|||
Db      3897 -----CCAGGTGTGTGTGAGGCAATTGGCTTCTCTGCTGCATCCCCCTTC 3942
Qy      584 -----ThrMetProAlaLeuAsnThrAlaPheSerThrGlnGlyArgAlaArgAspGlu 602
      |||      |||      |||      :|||
Db      3943 CAGCTCCCTCCCTGTCTCAGATCTGATACCAAGACAGAGGCCCGCGGCCAGCAGAGA 4002
Qy      602 rpAspThrMetPheLysIleProAsn 610
      |||      :|||
Db      4003 GCGACCAGCAGGCTCTCTCAGCCCTCT 4028
```

Search completed: July 18, 2005, 14:43:23
Job time : 8265 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 15, 2005, 13:09:22 ; Search time 924 Seconds
(without alignments)
4119.474 Million cell updates/sec

Title: US-10-736-868-2

Perfect score: 3291

Sequence: 1 MILFLFLLLGFCIAPLSA.....SKTRFVGNGAFDMPALGL 643

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/US10736868/runat_14072005_105509_2268/app_query.fasta_1.839
-DB=N_Geneseq_16Dec04 -QMT=fastcap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10736868 @CIGN 1.1 470 @runat_14072005_105509_2268 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_16Dec04:.*
1: Geneseq1980s:.*
2: Geneseq1990s:.*
3: Geneseq2000s:.*
4: Geneseq2001as:.*
5: Geneseq2001bs:.*
6: Geneseq2002as:.*
7: Geneseq2002bs:.*
8: Geneseq2003as:.*
9: Geneseq2003bs:.*
10: Geneseq2003cs:.*
11: Geneseq2003ds:.*
12: Geneseq2004as:.*
13: Geneseq2004bs:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	181	5.5	3631	4 ABL03905	Ab103905 Drosophil
2	180.5	5.5	7396	12 ADO09868	Ado09868 Human NOV
3	177	5.4	3543	4 AAH62819	AAH62819 Shrimp wh
4	177	5.4	305107	4 AAH62689	AAH62689 Shrimp wh
5	176.5	5.4	7396	6 ABO08505	Abt08505 Human nov

6	174.5	5.3	6432	5 AAS65903	Aas65903 DNA encod
7	173.5	5.3	7274	13 ADN60282	Adn60282 Human non
8	173.5	5.3	7516	4 AAK51987	Aak51987 Human pol
9	171	5.2	100990	12 ADJ94407	Adj94407 Versinia
10	169.5	5.2	6126	10 ADC13546	Adc13546 Human NOV
11	169.5	5.2	6151	6 ABZ11646	Abz11646 Human pol
12	169.5	5.2	6151	12 ADM44164	Adm44164 Novel hum
13	168	5.1	5439	4 ABL02411	Ab102411 Drosophil
14	168	5.1	5756	4 ABL03904	Ab103904 Drosophil
15	167.5	5.1	6968	4 ABL28617	Ab128617 Drosophil
16	165	5.0	3603	13 ADS49206	Ads49206 Bacterial
17	165	5.0	4755	13 ACN37449	Acn37449 Tumour-as
18	163.5	5.0	2932	9 ADB97572	Adb97572 Rat presy
19	163.5	5.0	6354	10 ADJ56352	Adj56352 Chicken c
20	163	5.0	2485	10 ADF90792	Adf90792 Human hep
21	163	5.0	4041	6 ABQ54663	Abq54663 Human ova
22	162.5	4.9	3257	13 ADS4956	Ads4956 Bacterial
23	162.5	4.9	5883	12 ADQ87119	Adq87119 Human tum
24	162.5	4.9	5883	13 ADQ84854	Adq84854 Human tum
25	162.5	4.9	5883	13 ACN40924	Acn40924 Tumour-as
26	162	4.9	3320	10 ADE85060	Ade85060 Farnesyl
27	162	4.9	6607	4 AAH98321	Aah98321 Rabbit ES
28	161.5	4.9	8486	4 AAK52971	Aak52971 Human pol
29	159	4.8	2680	2 AAQ84590	Aaq84590 AMML chro
30	159	4.8	2887	2 AAQ84589	Aaq84589 AMML chro
31	159	4.8	3388	6 ABL61748	Ab161748 Colon ade
32	159	4.8	3388	6 ABL66292	Ab166292 Lung canc
33	159	4.8	3388	6 ABL63718	Ab163718 Breast ca
34	159	4.8	3388	6 ABL64112	Ab164112 Breast ca
35	159	4.8	3388	6 ABL68053	Abt10826 Ovary can
36	159	4.8	3388	6 ABL10826	Abt10826 Human bre
37	159	4.8	4618	5 AAS74130	Aas74130 DNA encod
38	159	4.8	4936	3 AAC61196	Aac61196 Human cDN
39	159	4.8	5574	6 ABK35570	Abk35570 Gene MYH1
40	159	4.8	5835	9 ACA62125	ACA62125 Human smo
41	159	4.8	5937	9 ACA62126	ACA62126 Human smo
42	159	4.8	6861	8 ACF12909	Acf12909 Human cer
43	159	4.8	6861	12 ADK70431	Adk70431 Respirato
44	159	4.8	6861	12 ADN03744	Adn03744 Antipori
45	159	4.8	6900	8 ACF12910	Acf12910 Human cer

ALIGNMENTS

RESULT 1	
ABL03905	
ID	ABL03905 standard; cDNA; 3631 BP.
XX	
AC	ABL03905;
XX	
AC	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster expressed polynucleotide SEQ ID NO 6197.
XX	
KW	Drosophila; developmental biology; cell signalling; insecticide;
KW	pharmaceutical; gene; ss.
XX	
OS	Drosophila melanogaster.
XX	
FN	WO200171042-A2.
XX	
PD	27-SEP-2001.
XX	
PD	
PF	23-MAR-2001; 2001WO-US009231.
XX	
PR	23-MAR-2000; 2000US-0191637P.
PR	11-JUL-2000; 2000US-00614150.
XX	
PA	(PEKE) PE CORP NY.
XX	
FI	Venter JC, Adams M, Li PWD, Myers EW;
XX	
DR	WPI; 2001-656860/75.

QY 485 AlaLeuPheMetAspAspLysLeuGluAenThrLeuLysGlyArg----- 499
 Db 2633 GGACATCTCAGCGACGAGGAACTGCAAAACGACGGCGAGGAATGGAAGACGCGCCACACG 2692
 QY 500 -----GlnMetLeuThrAspGlu-----GlnLysGlyArg 509
 Db 2693 CGGAGGCGCCGAAGCCAAATCCAGGTGCTGCAGCAGAGGTTCGCTCAGGAATGAAG 2752
 QY 510 ThrArgValLysThrIleArgAlaLeuProArgLeuPheGlyAla----- 524
 Db 2753 AAGCAGACGAAAAGATCAAGCG-----CGCTGTGGTCCCGTTGGCTGGACGAG 2806
 QY 525 -----ProThrAlaLysAlaGluMetIleAspAlaLysVal 536
 Db 2807 AACGGCAACAAGTCCGAGCTATTCCAGCTATCTTGGCCACACCATCGACATGTAGGCC 2866
 QY 537 PheGlnAspIleGluGluArgProIleProProLeuPhePheGluProLysGlyArgHis 556
 Db 2867 TGTGGCAGGTGGAG-----CCACTCAGCCTAGAGCCTCC-GCCAGA--- 2907
 QY 557 ThrArgLeuArgTrpThrGlyAlaAenGluLysGluIleProGlyLeuGlySerArgPhe 576
 Db 2908 -----ACCGAGCGTCAGGATCAGATCAGCGCAGCAGCGCGTCCCGCTGCA 2952
 QY 577 IleLeuProSerLeuAspProThrMetProAlaLeuAenThrAlaPheSerThrGlnGly 596
 Db 2953 GCTGCAGCATCG-----TCTCATGCAGCAGCTGGTGGCGCTTACCCACGCAATCG 3003
 QY 597 ArgAlaArgAspGluTrpAsp---ThrMetPheLysIleProAsnAsnTrpAsnProGly 615
 Db 3004 GAATCGAAGACGTTCTCTGATCAACGAATATTAGTAGTACTGAA-----AACCCAAGC 3057
 QY 616 AspGluValGlyPheLysMetAsnSerLysThrLys 627
 Db 3058 GACCCCGAGGCGCAATCAACACGCTGCCCTCAAA 3093
 RESULT 2
 ADO09868
 ID ADO09868 standard; cDNA; 7396 BP.
 XX AC ADO09868;
 DT 01-JUL-2004 (first entry)
 XX Human NOV11 cDNA.
 KW human; ss; gene; NOVX; immunogen; vaccine; cancer; diabetes;
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease; asthma;
 KW allergy; emphysema; bronchitis; autoimmune disease;
 KW graft-versus-host disease; arthritis; scleroderma;
 KW systemic lupus erythematosus; bacterial infection; cystic fibrosis;
 KW coronary artery disease; stroke; hypertension; myocardial infarction;
 KW haemophilia; idiopathic thrombocytopenic purpura; hyperlipidaemia;
 KW obesity; cirrhosis; inflammatory bowel disease; Crohn's disease; ulcers;
 KW muscular dystrophy; myasthenia gravis; endometriosis; psoriasis;
 KW alopecia; uveitis; amyotrophic lateral sclerosis; osteoporosis;
 KW osteoarthritis; liver disease; epilepsy; multiple sclerosis; anxiety;
 KW pain; fertility; glomerulonephritis; polycystic kidney disease; SNP;
 KW single nucleotide polymorphism.
 XX Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT variation replace(5008,t)
 FT /*tag= a
 FT /note= "Single nucleotide polymorphism"
 FT variation replace(5012,g)
 FT /*tag= b
 FT /note= "Single nucleotide polymorphism"
 FT variation replace(6808,t)
 FT /*tag= c
 FT /note= "Single nucleotide polymorphism"

FT variation replace(7323,t)
 FT /*tag= d
 FT /note= "Single nucleotide polymorphism"
 PN US2004018970-A1.
 XX 29-JAN-2004.
 XX 27-MAR-2002; 2002US-00107782.
 XX 19-DEC-2000; 2000US-0256619P.
 PR 19-JAN-2001; 2001US-0262959P.
 PR 28-FEB-2001; 2001US-0272408P.
 PR 28-MAR-2001; 2001US-0279344P.
 PR 20-APR-2001; 2001US-0285189P.
 PR 26-JUL-2001; 2001US-0308039P.
 PR 09-AUG-2001; 2001US-0311266P.
 PR 19-DEC-2001; 2001US-00028248.
 XX (SHIM/) SHIMKETS R A.
 PA (PATT/) PATTURAJAN M.
 PA (VERN/) VERNET C A M.
 PA (CASM/) CASMAN S J.
 PA (MALY/) MALYANKAR U M.
 PA (SHEN/) SHENOY S G.
 PA (SPYT/) SPYTEK K A.
 PA (GANG/) GANGOLLI E A.
 PA (MILL/) MILLER C E.
 PA (BOLD/) BOLDOG F L.
 PA (LILL/) LI L.
 PA (TAUP/) TAUPIER R J.
 PA (KEKU/) KEKUDA R.
 PA (SMIT/) SMITHSON G.
 PA (ZERH/) ZERHUSEN B D.
 PA (LIXX/) LIU X.
 PA (COLM/) COLMAN S D.
 PA (TCHE/) TCHERNEV V T.
 PA (SIJJ/) SI J.
 PA (EDIN/) EDINGER S R.
 PA (STON/) STONE D J.
 PA (SCIO/) SCIORE P.
 PA (MILL/) MILLET I.
 PA (ROTH/) ROTHENBERG M E.
 XX Shimkets RA, Patturajan M, Vernet CAM, Casman SJ, Malyankar UM;
 PI Shenoy SG, Spytek KA, Gangolli EA, Miller CE, Boldog FL, Li L;
 PI Taupier RJ, Kekuda R, Smithson G, Zerhuseen BD, Liu X, Colman SD;
 PI Tchernev VT, Si J, Edinger SR, Stone DJ, Sciore P, Millet I;
 PI Rothenberg ME;
 XX WPI: 2004-122080/12
 DR P-PSDB; ADO09869, ADO09979.
 XX New NOVX polypeptides and nucleic acid molecules, useful for diagnosing,
 PT preventing or treating NOVX-associated disorders e.g. cancer, diabetes,
 PT Alzheimer's disease, stroke, arthritis, hypertension or myocardial
 PT infarction.
 XX Claim 8; SEQ ID NO 35; 240pp; English.
 XX The invention relates to an isolated NOVX polypeptide. A therapeutic,
 CC i.e. the NOVX polypeptide, nucleic acid and antibody, is useful for
 CC manufacturing a medicament for treating a syndrome associated with a
 CC human disease, e.g. a NOVX-associated disorder. The NOVX polypeptides can
 CC be used as immunogens or as vaccines. The NOVX polypeptide, nucleic acid
 CC or antibody is useful for diagnosing, treating or preventing a NOVX-
 CC associated disorder, e.g. cancer, diabetes, Alzheimer's disease,
 CC Parkinson's disease, Huntington's disease, asthma, allergies, emphysema,
 CC bronchitis, autoimmune disease, graft-versus-host disease, arthritis,
 CC scleroderma, systemic lupus erythematosus, bacterial infections, cystic
 CC fibrosis, coronary artery disease, stroke, hypertension, myocardial
 CC infarction, haemophilia, idiopathic thrombocytopenic purpura,
 CC hyperlipidaemia, obesity, cirrhosis, inflammatory bowel disease, Crohn's

QY 462 ----- 462
 Db 5815 CCGAAACTGACGCGAGCTGGAGGAGCGCACTGAGAGCGCGATGCCATGAACCGCA 5874
 QY 463 -----Thrg 464
 Db 5875 AGTCAGCTCCCTAAAGAACAGCTCAGGCGGGGACCTGCGTTGTGTGTCGCCCGCG 5934
 QY 464 luTrpLys-----IleAlaLysGluSerA 472
 Db 5935 AATGCCCGAAGCGCGGGATGGTCCACGAGAGGTAGATGCAAGCGGATGG 5994
 QY 472 rgValLeuThrAsnLeuLysAsnProSerLeuAlaLeuPheMetAspLysL 492
 Db 5995 GGCTGAGGCAAACTGCGGAATAAGCTCTCTCTGCGAGCTCAGATGGATGACAGA 6054
 QY 492 euGluAsnThrLeuLysGlyArgGlnMetLeuThrAspGluGlnLysGlyArgThrArgV 512
 Db 6055 CAGACACACAGCTCCCTTCCAG-----ACCCGCGAG----- 6089
 QY 512 alLysThrIleArgAlaLeuProArgLeuPheGlyAla-----ProThrA 527
 Db 6090 -----CAGCCCTCTCCCACTCTTGGGACTGTGTGAACATGCTCTCTCTG 6138
 QY 527 laLysAlaGluMetIleAspAlaLysValPheGlnAspIleGluArgProIleProp 547
 Db 6139 CCCTC-----CGCCCGCGTCCCG 6156
 QY 547 roLeuPhePheGluProLysGlyArgHisThrArgLeuArgTrpThrGlyAlaAsnGluL 567
 Db 6157 CATCCCGTTTCCCT----- 6170
 QY 567 ysGluIleProGlyLeuGlySerArgPheIleLeuProSerLeuAspPro-----T 584
 Db 6171 -----CCAGGTGTGTGGGCAATTGGCTTCTCTGCTGATCCCTTCCAGCTC 6222
 QY 584 hrMetProAlaLeuAsnThrAlaPheSerThrGlnGlyArgAlaArgAspGluTrpAspT 604
 Db 6223 CTCTCCCTGCTCAGAATCTGATACCAAGAGACAGCGCGCGCCGAGCAGAGCGGACC 6282
 QY 604 hrMetPheLysIleProAsn 610
 Db 6283 AGCAGGCTCTCAGCCCTCT 6302
 RESULT 6
 AAS65903
 ID AAS65903 standard; cDNA; 6432 BP.
 AC AAS65903;
 DT 13-FEB-2002 (first entry)
 XX DNA encoding novel human diagnostic protein #1707.
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX Homo sapiens.
 OS
 XX WO200175067-A2.
 FN
 XX 11-OCT-2001.
 PD
 XX 30-MAR-2001; 2001WO-US008631.
 XX
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX
 XX (HYSE-) HYSEQ INC.
 XX
 FI Drmanac RT, Liu C, Tang YT;
 XX

DR WPI; 2001-639362/73.
 DR P-PSDB; ABG01716.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 1; SEQ ID NO 1707; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 6432 BP; 1682 A; 1699 C; 2071 G; 980 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0.00073 Length: 6432
 Score: 174.50 Matches: 167
 Percent Similarity: 35.04% Conservative: 107
 Best Local Similarity: 21.36% Mismatches: 278
 Query Match: 5.30% Indels: 232
 DB: 5 Gaps: 32
 US-10-736-868-2 (1-643) x AAS65903 (1-6432)
 QY 12 GlyPheCysIleAlaProLeuSer-----AlaGlnSerProSerThrSerAsp----- 27
 Db 4106 GGCTTCTCAGCCAGTCCGACGACGAGTCCAGCAAGCTCACCAGGACTTCTCCGGCGCTGG 4165
 QY 28 AlaProGly-----AlaLeuLeuSerSerLeuValGlyLysSerHisGlnLysLeuPro 45
 Db 4166 AGTCCAGCTGCGAGGACACTCAGGAGCTG-CTGCAGGAGGAGAGACCGCGCAGAGCTGAGC 4224
 QY 46 LeuAlaProSerMetGluAlaLeuGlu----- 54
 Db 4225 CTGAGCACCAGCTCAAGCAGGCTGAGGAGCAGAGAAGAAATTCCTTCCGGGAGCAGCTGGAG 4284
 QY 55 -----LeuMetGlyValGlnPhe 60
 Db 4285 GAG 4344
 QY 61 ValAspAlaLeuIleLysLysGlyGlnMetGluMetAlaLysGlyAlaPheLysThrGln 80
 Db 4345 GCCGAC-----ATGAAAAAGAGATGGAGGACAGTGTGGGGTGTCTCGAAACTGCT 4395
 QY 81 LeuGluValLeuGluLysValHisProAsp----- 90
 Db 4396 GAGGAGGTGAAGAGGAAGCTCCAGAGGAGCCTGTGAGGCGCTGTGAGCCAGCGCAGGAGAG 4455
 QY 91 -----GlnPheAspLysTyrlLysLysLysLys-----ValAsp 101
 Db 4456 AAGTGGCGCGCTACGACAAAGCTCGGAGAGACCGAGAGCGGGCTGCGAGCAGGAGCTGGAC 4515
 QY 102 AspLeuAlaAlaAspAlaValMetGlnGlnAlaGluMetAlaLysLeuGlnProLysSer 121

KW sugar transporter; tumorigenesis; antiangiogenic; cytostatic;
 KW cerebroprotective; vasotropic; antiinfertility; cardiant;
 KW antibody therapy; antisense therapy; RNA interference therapy;
 KW RNAi therapy; cancer; stroke; infertility; heart disease; human; gene;
 KW ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 1..5883

FT /tag= a

FT /product= "non-muscle myosin heavy chain 9"

XX WO2004039955-A2.

XX 13-MAY-2004.

XX 29-OCT-2003; 2003WO-US034281.

XX 29-OCT-2002; 2002US-0421989P.

XX 17-OCT-2003; 2003US-0512251P.

XX (RIGE-) RIGEL PHARM INC.

XX Lorens JB, Atchison RE, Frieria A, Holland S;

XX WPI; 2004-376181/35.

XX P-PSDB; ADN60283.

XX Identifying a compound that modulates angiogenesis or tumorigenesis,
 XX useful in diagnosing and treating angiogenesis, cancer, stroke,
 XX infertility and heart disease, comprises contacting the compound with
 XX angiogenesis polypeptide.

XX Disclosure; Page 71-72; 105pp; English.

XX The present invention describes a method for identifying a compound that
 CC modulates angiogenesis or tumorigenesis. The method comprises: (a)
 CC contacting the compound with angiogenesis polypeptide, e.g. Axi, tubulin
 CC cofactor D, transglutaminase 2, cytosine deaminase, peptidase M41
 CC (paraplegin), CD13 aminopeptidase, PRK-1, zip kinase, Gas6, SRml60, non-
 CC muscle myosin heavy chain, calmodulin 2, novel symporter, novel
 CC seaphorin, novel zinc finger helicase (FLJ22611), plexin-A2,
 CC deoxyribidylate deaminase or novel sugar transporter; (b) determining the
 CC functional effector of the compound upon the angiogenesis polypeptide or
 CC the physical effect of the compound upon the target polypeptide or its
 CC fragment or inactive variant; and (c) determining the chemical or
 CC phenotypic effect of the compound upon a cell comprising the target
 CC polypeptide or its fragment or inactive variant, thus identifying a
 CC compound that modulates cell cycle arrest. Also described is a method of
 CC modulating angiogenesis in a subject. The angiogenesis or tumorigenesis
 CC modulating compound has antiangiogenic, cytostatic, cerebroprotective,
 CC vasotropic, antiinfertility and cardiant activities, and can be used in
 CC antibody, antisense and RNA interference (RNAi) therapies. The method is
 CC useful in identifying a compound that modulates angiogenesis. The methods
 CC and compounds or compositions are useful in diagnosing and treating
 CC angiogenesis, cancer, stroke, infertility and heart disease. The present
 CC sequence encodes a human non-muscle myosin heavy chain 9 protein, which
 CC is used in the exemplification of the present invention.

XX SQ Sequence 7274 BP; 1861 A; 1977 C; 2247 G; 1189 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0.00103	Length:	7274
Scores:	173.50	Matches:	178
Percent Similarity:	30.71%	Conservative:	91
Best Local Similarity:	20.32%	Mismatches:	263
Query Match:	5.27%	Indels:	346
DB:	13	Gaps:	35

US-10-736-868-2 (1-643) x ADN60282 (1-7274)

Qy 12 GlyPheCysIleAlaProLeuSer-----AlaGlnSerProSerThrSerAsp----- 27

DB	Sequence	Position
Db	3860 GGCTTCTCAGCCAGTCCGACGACGAGTCCAGCAAGCTCACCAGGACTTCTCCGGCTGG	3919
Qy	28 AlaProGly-----AlaLeuLeuSerSerLeuValGlyLysSerHisGlnLysLeuPro	45
Db	3920 AGTCCCACTGAGGACACTCAGGAGCTG-CTGCAGGAGGAGAACCGGAGAAGCTGAGC	3978
Qy	46 LeuAlaProSerMetGluAlaLeuGlu-----	54
Db	3979 CTGAGCACCAGCTCAAGCAGGTGGAGGACGAGAGAATTCTCTCCGGGAGCAGCTGGAG	4038
Qy	55 -----LeuMetGlyValGlnPhe	60
Db	4039 GAGGAGGAGGAGGCCAAGCACAACCTGGAGAAGCAGATCGCCACCTCCATCCCAAGGTG	4098
Qy	61 ValAspAlaLeuIleLysLysGlyGlnMetGluMetAlaLysGlyAlaPheLysThrGln	80
Db	4099 GCCGAC-----ATGAAAAGAGATGGAGGACAGTGTGGGTGCTTGGAAAAGTGT	4149
Qy	81 LeuGluValLeuGluLysValHisProAsp-----	90
Db	4150 GAGGAGGTGAAGAGGAGCTCCAGAGGACCTGGAGGCGCTGAGCCAGCGGCACGAGGAG	4209
Qy	91 -----GlnPheAspLysTyrlsLysLysLys-----ValAsp	101
Db	4210 AAGGTGGCGCGCTACGACAGCTGAGAGAACCAAGACGCGCTGCAGCAGGAGCTGGAC	4269
Qy	102 AspleuAlaAlaAspAlaValMetGlnGlnAlaGluMetAlaLysLeuGlnProLysSer	121
Db	4270 GACCTGTGTGGACCTGGACACACGCGCAGAGCGCGTGCACCTGGAGAAGAAGCAG	4329
Qy	122 GlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIleProIleGly-----	137
Db	4330 AAGAAG-----TTTGACCACTCTCTGGCGGAGGAGACCATCTCTGCCAAGTATGCA	4383
Qy	138 -----SerSer	139
Db	4384 GAGGAGCGCGACCGGCTGAGCGGAGCGCCGAGAGAGGAGACCAAGGCTCTGTCTGCTG	4443
Qy	140 IleArgGlyLeuGluAspAlaIleArgThrGlnArgAspMetGluAsnThrAspProSer	159
Db	4444 GCCCGGCGCTGGAGGAGGAGCATGTGAGCAGAGCGGAGCTGGAGCGGCTCAAC-----	4497
Qy	160 GluGlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeuProGlyLeuVal	179
Db	4498 -----AAGCAGTTCCGACGGAG-----ATG	4518
Qy	180 AlaAsnMetIleAlaGlyLysAsnProPhe-----LysMetProGlnGlnMetArgLysAla	198
Db	4519 GAGGACCTTATGAGTCCAGGATGATGTGGGCAAGAGTGTCCACGAGCTGGAGAAGTC-	4577
Qy	199 GlnAlaAlaProSerSerValPheGlnGlnAlaLeuAlaGlnArgAlaMetLeuGlyLys	218
Db	4578 CAACGGGCGCT-----AGAGCAGCAGGTGGAGGA	4607
Qy	219 Asn-----AlaProValAlaGlyLysArgGlyGluGluGlnArgMetMetMetAsn	235
Db	4608 GATGAAGACGCGAGCTGGAGAGCTGGAGGAGCTGCAGCGCCACCGAAGATGCCAAGCT	4667
Qy	236 -ArgValAspGlnArgMetGln-----GlnArgGluLeuGlnGlu	249
Db	4668 GCGGTGGAGGTCAACCTGCGAGGCCCATGAAGGCCAGTTCGAGGGGACCTTCGAGGGCGG	4727
Qy	249 uAspGluAspAspAspAspLeuGluAspGluAspValProArgArgArgSerSerAspGln	269
Db	4728 GGACGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	4787
Qy	269 yGluProGlnSerGluAlaGluHisGlnArgArgAspLeuAlaArgArgLeuLysSerSe	289
Db	4788 AGAGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	4847
Qy	289 rProArgLysLysGluLeu-----	295

Db 4848 GATGCACCTGAGGAGCTGGAGCGCACATCGACTCGGCCAACAAAGAACCGGACGACG 4907
 Qy 296 -----LeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrGlnArgMetArgAs 312
 Db 4908 CATCAAAACAGCTCGGAAGCTGCAGGCGCCAGATGAGAGCTGATGCGAGCTGGATGA 4967
 Qy 312 pSerProLeuSerIysArgArgProLeuAlaMetAsnAspGluAspGluSerAlaPheAr 332
 Db 4968 CACCCGCGCTCTCGTGAGAGATCTCGGCCAGGCCAAAGAGAACACGAGAAAGAGCTGAA 5027
 Qy 332 gAlaMetGluAlaArgAlaLysLeuAspGlnLysSerGlnLeuValLeuGlyLeuHisG1 352
 Db 5028 GAGCATGGAGGCC-----GAGATGATCCATTGTCAGGAGGAATGGCAGC 5072
 Qy 352 yPheGlyGluSerAspAspGluAspGluGluAspGluAsnLeuLeaPheSerG1 372
 Db 5073 CGCGGAGCGTGCCAAAGCGCCAGCGCCAGCAGGAGCGGGATGAGCTGACGAGATCGC 5132
 Qy 372 uAnSerPheArgArgAlaProLeuArgLeuSer----- 383
 Db 5133 CAACAGCAGCGGCAAAAGAGCCCTGGCGTTAGAGGAGAACCGCGCTGTGAGGCCCGCAT 5192
 Qy 384 -SerGlyPheValGluLysLeuLysSer-----AsnAspG1 395
 Db 5193 CGCCAGCTGGAGGAGAGCTGGAGGAGGAGCAGGCGCAACACGGAGCTGATCAACGACCG 5252
 Qy 395 uLeuLysSerAla----- 399
 Db 5253 GCTGAAGAAGGCCAAACCTGCAGATCGACACAGCAGATCAACACCGACTGAACCTGGAGCGCAG 5312
 Qy 400 -----LeuAspArg-----lleLys 404
 Db 5313 CCACGCCCAAGAACGAGAAATGCTCGGACGAGCTGGAAAGCGCCAGAACGAGGAGCTTAA 5372
 Qy 404 sTyrArgValAspAspValGlu-----LysTyrLeuAla-ProLysPro- 418
 Db 5373 GGTCAAGCTCAGGAGATGAGGGCACTGTCAAGTCCAAGTACAAGGCTCCATCACC GC 5432
 Qy 419 -----MetGluPheAsnProLysPro----- 425
 Db 5433 CCTCGAGGCCAAGATTGCACAGCTGGAGGAGCAGCTGGCAACAGCAGAACGAGGCGCCA 5492
 Qy 426 --GlnProGlyTyr----- 429
 Db 5493 GGCAGCCTGCAAAACAGGTGCTGCGACCGAGAACAGCTGAAGATGTGCTGTGACGAT 5552
 Qy 430 -----PheAlaProArgLysIlePr-oThrArgProArgLysMetLeuProL 445
 Db 5553 GGATGACGAGCGGAGGAACCGCCGAGCAGTACAGGACCGCCGACAGGCAATCTACCCG 5612
 Qy 445 euLeuIleGlySerAspProLysValGlnGluIleArgArgHisProSer----- 462
 Db 5613 CCTGAAGCAGCTCAAGCGGCGAGCTGGAGGAGCGCG--AAGAGGAGGCCCGAGCGGCCAA 5669
 Qy 462 ----- 462
 Db 5670 CGCCTCCCGCGGAAACTGCAGCGCGAGCTGGAGGACGCCCACTGAGACGGCGCATGCCAT 5729
 Qy 462 ----- 462
 Db 5730 GAACCGGAAGTCAGCTCCCTAAGAACAGCTCAGCGCGGGGACCTGCGGTTGTGCT 5789
 Qy 463 -----ThrGluTrpLys-----IleAlaL 469
 Db 5790 GCCCGCGGAATGCCCGGAAAGCGCGGGATGGCTCCGACGAAGAGGTAGATGCCAA 5849
 Qy 469 yGluSerArgValLeuThrAsnLeuLysAsnAsnProSerLeuAlaLeuPheMetA 489
 Db 5850 AGCGGATGGGCTGAGCGCAACCTCGCGAATAGCCTCTCTCTCAGCCTGAGATGG 5909
 Qy 489 sPaspLysLeuGluAsnThrLeuLysGlyArgGlnMetLeuThrAspGluGlnLysGlyA 509
 Db 5910 ATGACACAGACAGACACACAGCCTCCCTTCCAG-----ACCCGCGAG----- 5953

Qy 509 rgThrArgValLysThrIleArgAlaLeuProArgLeuPheGlyAla----- 524
 Db 5954 -----CAGCGCTCTCCCCACCTTCTTGGGACTGTGTGAACATGC 5993
 Qy 525 --ProThrAlaLysAlaGluMetIleAspAlaLysValPheGlnAspIleGluGluArgP 544
 Db 5994 CTCCTCTGCCCTC-----CGCC 6011
 Qy 544 rolleProProLeuPhePheGluProLysGlyArgHisThrArgLeuArgTrpThrGlyA 564
 Db 6012 CCGTCCCCCATCCCGTTCCCT----- 6034
 Qy 564 laAsnGluLysGluIleProGlyLeuGlySerArgPheIleLeuProSerLeuAspPro- 583
 Db 6035 -----CCAGGTGTTGTTGAGGGCATTTGGCTTCTCTGCTGCATCCCC 6077
 Qy 584 -----ThrMetProAlaLeuAsnThrAlaPheSerThrGlnGly-ArgAlaArgAsp 600
 Db 6078 TTCCAGCTCCTCCCTGCTCAGAACTCATACCAAGAGACAGGCGCCGCGGCCAGGCA 6137
 Qy 601 GluTrpAspThrMetPheLysIleProAsn----- 610
 Db 6138 GAGAGCGACCGCAGCGCTCCTCAGCCTCTCTTCCAAAAAGCACAAGATGTTGAGCGCA 6197
 Qy 611 ----AsnTrpAsnProGlyAspGluValGlyPheLysMetAsn 623
 Db 6198 GCAGGCGAGCGCCCCCGGGAGGGCGCAGAGTTTCTATGAAT 6239
 RESULT 8
 AAK51987
 ID AAK51987 standard; cDNA; 7516 BP.
 XX AAK51987;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 532.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157190-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WO-US004098.
 XX
 PR 03-FEB-2000; 2000US-00496914.
 PR 27-APR-2000; 2000US-00560875.
 PR 20-JUN-2000; 2000US-00598075.
 PR 19-JUL-2000; 2000US-00620325.
 PR 01-SEP-2000; 2000US-00654936.
 PR 15-SEP-2000; 2000US-00663561.
 PR 20-OCT-2000; 2000US-00693325.
 PR 30-NOV-2000; 2000US-00728422.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
 PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 XX
 DR WPI; 2001-476283/51.
 DR P-PSDB; AAM78854.
 XX
 PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
 PT in diagnosis and gene therapy.
 XX

Qy	12	GlyPheCysIleAlaProLeuSer-----AlaGlnSerProSerThrSerAsp-----	27
Db	4094	GGCTTCTCAGCGATCCGACAGAGTCCAGCAGACTCACCAGAGACTTCTCCCGCTGG	4153
Qy	28	AlaProGly-----AlaLeuLeuSerSerLeuValGlyLysSerHisGlnLeuLeuPro	45
Db	4154	AGTCCCGAGCTCAGGACACTCAGAGCTG-CTGCAGGAGGAGAACCGGACAGACTGAC	4212
Qy	46	LeuAlaProSerMetGluAlaLeuGlu-----	54
Db	4213	CTGAGCCACCAAGCTCAAGCAGGTGGAGGACGAGAGAAGATTCTCTCCGGAGCAGCTGGAG	4272
Qy	55	-----LeuMetGlyValGlnPhe	60
Db	4273	GAGGAGGAGGGCCCAAGCACACTGGAGAGCAGATCGCCACCCCTCCATGCCCAAGTGT	4332
Qy	61	ValAspAlaLeuIleLysLysGlyGlnMetGluMetAlaLysGlyAlaPheLysThrGln	80
Db	4333	GCCGAC-----ATCAAAAAGAAGTGGAGGACAGTGTGGGGTGCTTGGAAACTGCT	4383
Qy	81	LeuGluValLeuGluLysValHisProAsp-----	90
Db	4384	GAGGAGGTGAAGAGGAGACTCCAGAGGACTTGGAGGGCCTTGAGCCAGCGGCACGAGGAG	4443
Qy	91	-----GlnPheAspLysTyLysLysLeuLys-----ValAsp	101
Db	4444	AAGGTGCCCGCTACGACAGCTGGAGAGACCAAGACGCGGTGCAGCAGGAGCTGCAC	4503
Qy	102	AspLeuAlaAlaAspAlaValMetGlnGlnAlaGluMetAlaLysLeuGlnProLysSer	121
Db	4504	GACCTGTGTGGACCTTGGACCACAGCGCGAGCGCGTGCACTGGAGAGAAGACAG	4563
Qy	122	GlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIleProIleGly-----	137
Db	4564	AAGNAG-----TTTGTACCAGCTCTCGCGGAGGAGAGAAGCACTCTCTGCCAAGTATGCA	4617
Qy	138	-----SerSer	139
Db	4618	GAGGAGCGCACCGGGCTGAGCGCGGCGCCGAGAGAAGGAGACCAAGGCTCTGCTCGCTG	4677
Qy	140	IleArgGlyLeuGluAspAlaIleArgThrGlnArgAspMetGluAsnThrAppProSer	159
Db	4678	GCCCGGGCCCTTGGAGGAAGCCATGGAGCAAGAACGGGAGCTGGGAGCGGCTCAAC-	4731
Qy	160	GluGlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeuProGlyLeuVal	179

```
Db 5727 GGCAGCCTGCAACACAGTGGCTCGGACCGAGAGAACTGAAGATGTGCTGCTGCAGGT 5786
Qy 430 -----PheAlaProArgLysIleProThrArgProArgLysMetLeuProL 445
Db 5787 GGATGACGAGCGGAGGAACCGCAGCAGTAGTACAGGACGAGCGGACACAGGCATCTACCCG 5846
Qy 445 euLeuIleGlySerAspProLysValGlnGluGluIleArgArgHisProSer----- 462
Db 5847 CTGAAGCAGCTCAAGCGGAGCTGGAGGAGCGG-----AAGAGAGGCGCCAGCGGGCCAA 5903
Qy 462 ----- 462
Db 5904 CGCTCCCGCGGAAACTGCAGCGCAGCTGGAGGACGCCACTGAGAGCGCCGATGCCAT 5963
Qy 462 ----- 462
Db 5964 GAACCGGAAAGTCAGTCCCTTAAGAACAAAGCTCAGGCGGGGACCTGCCTGTTGTCGT 6023
Qy 463 -----ThrGluTrpLys----- 469
Db 6024 GCGCCGCGGAATGCGCGGAAAGCGCGGGATGGCTCCGACGAGAGGATAGATGCGAA 6083
Qy 469 ySgluSerArgValLeuThrAsnLeuLysAsnAsnProSerLeuAlaAlaLeuPheMetA 489
Db 6084 AGCGGATGGGCTGAGGCGCAACTGCGGATAGCTCTCTCTCCAGCCCTGAGATGG 6143
Qy 489 spAspLysLeuGluAsnThrLeuLysGlyArgGlnMetLeuThrAspGluGlnLysGlyA 509
Db 6144 ATGGACAGACAGACACACAGCCCTCCCTTCCAG-----ACCCGCGAG----- 6187
Qy 509 rgThrArgValLysThrIleArgAlaLeuProArgLeuPheGlyAla----- 524
Db 6188 -----CACGCTCTCCCACTTCTTGGGACTGCTGTGAACATGC 6227
Qy 525 --ProThrAlaLysAlaGluMetIleAspAlaLysValPheGlnAspIleGluGluArgP 544
Db 6228 CTCTCTCTGCGCTC----- 6245
Qy 544 rolleProProLeuPhePheGluProLysGlyArgHisThrArgLeuArgTrpThrGlyA 564
Db 6246 CCGTCCCGCCCATCCGTTCCCT----- 6268
Qy 564 laAsnGluLysGluIleProGlyLeuGlySerArgPheIleLeuProSerLeuAspPro- 583
Db 6269 -----CCAGGTTGTTGAGGCGATTGGCTTCCCTGCTGCTGATCCCC 6311
Qy 584 -----ThrMetProAlaLeuAsnThrAlaPheSerThrGlnGly-ArgAlaArgAsp 600
Db 6312 TTCCAGCTCCCTCCCTGCTCAGAACTCTGATACCAAGAGACAGGCGCGGCGCCAGGCA 6371
Qy 601 GluTrpAspThrMetPheLysIleProAsn----- 610
Db 6372 GAGAGCGACGACGAGGCTCTCTCAGCCCTCTCTTGGCCAAAAGACACAGATGTTGAGCGCA 6431
Qy 611 ---AsnTrpAsnProGlyAspGluValGlyPheLysMetAsn 623
Db 6432 GCAGGCGAGGCGCGCGGAGGGGCGCAGAGTTTCTATGAAT 6473
RESULT 9
ID ADJ94407
XX ADJ94407 standard; DNA; 100990 BP.
XX AC ADJ94407;
XX CC
XX DT 03-JUN-2004 (first entry)
XX DE Yersinia pestis plasmid pMT1.
XX OS Yersinia pestis
XX KW Bubonic plague; ds; plasmid; cyclic; circular; pMT1; capsular protein.
XX OS Yersinia pestis.
XX PN US6706522-B1.
```

```
XX 16-MAR-2004.
XX PD
XX PF 30-SEP-1999; 99US-00409800.
XX PR 30-SEP-1999; 99US-00409800.
XX PA (WISC ) WISCONSIN ALUMNI RES FOUND.
XX PI Blattner FR, Burland V, Rose DJ, Mayhew GF, Perna N, Perry RD;
XX PI Straley SC, Fetherston JD, Lindler LE, Plano GV;
XX DR WPI; 2004-236724/22.
XX PT New recombinant DNA construct comprising an open reading frame placed
XX PT under the control of a non-native promoter, useful for preparing a
XX PT composition for diagnosing, treating or preventing infection caused by
XX PT Yersinia pestis.
XX PS Example; SEQ ID NO 2; 90pp; English.
XX CC The invention relates to a recombinant DNA construct comprising an open
XX CC reading frame placed under the control of a non-native promoter. The open
XX CC reading frame comprises bp 2389-2826 of the Yersinia pestis plasmid
XX CC pPCPI, representing the pesticin immunity protein, a bacteriocin. Also
XX CC disclosed are the sequences of the Yersinia plasmids pCD1 (responsible
XX CC for the complex virulence response termed LCR, low calcium response) and
XX CC pMT1 (encoding capsular proteins and a murine toxin). The recombinant DNA
XX CC construct is useful for preparing a composition for diagnosing, treating
XX CC or preventing infection caused by Yersinia pestis (e.g. bubonic plague).
XX CC The present sequence is the pMT1 plasmid.
XX SQ Sequence 100990 BP; 26256 A; 24506 C; 26153 G; 24075 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 0.0423 Length: 100990
Score: 171.00 Matches: 142
Percent Similarity: 33.99% Conservative: 100
Best Local Similarity: 19.94% Mismatches: 260
Query Match: 5.20% Indels: 210
DB: 12 Gaps: 33

US-10-736-868-2 (1-643) x ADJ94407 (1-100990)
Qy 17 ProLeuSerAlaGlnSerProSer---ThrSerAspAlaProGlyAlaLeuLeuSerSer 35
Db 85225 CCAGTGTGTGTAACATCCCATCTATACCGGACGACGCATCACCAACCCCTCATGAACGCA 85284
Qy 36 LeuValGly----- 38
Db 85285 GTACGTGGTGTTCGACCATGAAGTAGCCCATATCTCTTTACTGATCGAAAGTCGCG 85344
Qy 39 -----LysSerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGluLeu 55
Db 85345 ATGAAGATGCGCGAGAGAGAAAGGCTCGCTCTACCGGGCTTTGGAAACGCGCTGAAA--- 85401
Qy 56 MetGlyValGlnPheValAspAlaLeuIleLysLysGlyGlnMetGluMetAlaLysGly 75
Db 85402 -----GACGTTTTTTATTGAACGAGAAATGGGACAGGTATTTCACACGGA 85443
Qy 76 AlaPheLysThrGlnLeuGluValLeuGluLysValHisProAspGlnPheAspLysTyr 95
Db 85444 ACCGTGCGTAATTGTAGCCACACAGAACCTGGTG-----ATCGCAATAATAC 85491
Qy 96 LysLysLeuLysValAspAspLeuAlaAlaAspAlaValMetGlnGlnAlaGluMet--- 114
Db 85492 TTCAAGGCGCAAGTGTGAGAGCGGGTTTCAATCTGCACGCGCAACACGCGTGAATTGTTTC 85551
Qy 115 -----AlaLysLeuGlnProLysSerGlyAsnAla---PheIleAsp 127
Db 85552 CTGAAATTTCTTCCTTTGCCCGGTCGTCGCGGCTGGGATGGCCAAAGCCCTTTTCATCGAC 85611
Qy 128 MetLeuAsnGlyAsnGlyIleProIleGlySerIle-----ArgGly 142
```

```
Db 85612 TTTATCGAAGAACACTGGCACCCTCATCGATAAGCCCTGTAGCCTTGTCAAAGAGCATGGT 85671
QY 143 LeuGluAspAlaIleArgThrGlnArgAspMetGluAsnThrAspProSerGluGlnIle 162
Db 85672 ATCGAGCTGGCGCTT-----CCAAATATGTGGAACACAGAGACTGCGTAAGGTC 85722
QY 163 AlaLysAlaValMetAspLysPheGlnThrGlnIleLeuProGlyLeuValAlaAsnMet 182
Db 85723 GCAGCTGCTATCGCCCAATATCGCAGGAT----- 85752
QY 183 IleAlaGlyAsnProPheLysMetProGlnGlnMetArgLysAlaGlnAlaPro 202
Db 85753 ATGAAAGACAGCCAGAGGCGCAATTACCT--GAGCTTAAATCATCTGCCAAAAGCGC 85809
QY 203 SerSerValPheGlnGlnAlaLeuAlaGlnArgAlaMetGluLysAsnAlaProVal 222
Db 85810 TCTAAGACGAAGACAGTCC-----GAGAAACACACAGAG 85845
QY 223 AlaGlyGlyArgGlyGluGlnArgMetMetMetAsnArgValAspGlnArgMetGln 242
Db 85846 TCAGCGCAGCATCGCTCTCATAGTGAATCTGCACCAAGCGAAACCAAGGCGAAGCAGAC 85905
QY 243 GlnArgGluLeuGlnGluAsp---GluAspAspAspLeuGluAspGluAspVal 261
Db 85906 GACAAGGAAGACGAAGAGATGATCGCTCAGAAGAGGAAGATTCTGGGGATTCTGATTG 85965
QY 262 ProArgArgArgSerSerAspGlyGluProGlnSerGluAlaGluHisGlnArgAsp 281
Db 85966 CCT-----GATTCAATAGATAGGACTTACCTATATACATGAT----- 86001
QY 282 LeuAlaArgArgLeuLysSerSerProArgLeuLysGluLeuGlnAsnAlaGluVal 301
Db 86002 -----AAAGAAATTAGTGATACAGAAAGTAAA 86028
QY 302 GlnSerLeuLeuSerTyrGlnArgMetArgAspSerProLeuSerLysArgArgProLeu 321
Db 86029 TATACAGAAGCAGCGGCGAAGATGAGTCAGGAGACACCCAGATCCGATCGATCGCGCATG 86088
QY 322 AlaMetAsnAsp---GluAspGluSerAlaPheArgAlaMetGluAla----- 336
Db 86089 GAATCAGGTGATCTTGACACGACGAGGTGTAGTATGGCAGTCGCGCTCCCAACGCTGGT 86148
QY 337 -----ArgAlaLysLeuAspGlnLysSerGlnLeuValLeuGlyLeuHisGlyPhe 353
Db 86149 GATGCGATTCCGAAGACCGCATGACTCCGATGTTATGGCTTGGCGCGCTGGCGAT 86208
QY 354 GlyGluSerAsp-----AspAspGluAsp 361
Db 86209 GGGGATAGTGATAACGCGCAAGACTCCGATGCGGTCGTGAGGAGTCCGAGGAGAAGAC 86268
QY 362 GluGluAspGluAsnLeuIleAspProSerGluAsnSerPheArg----- 376
Db 86269 GAGCGGAAGAGAGCTGCCGATCACAATGACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 86328
QY 377 ---ArgAlaProLeuArgLeuSerSerGlyPheVal-----GluLysLeuLysSer 392
Db 86329 GCGGAAGCGCCTGAAGACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 86388
QY 393 AsnAspGluLeuLysSerAlaLeuAspArgLysTyrArgValAspAspVal----- 410
Db 86389 GAGGACGCACTCAAG---GCGCTCAGCAGATGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 86445
QY 411 -----GluLysTyrLeuAlaProLysProMetGluPheAsn 422
Db 86446 GAGCAGCTGTGGCCACCATCAGCAAGAGCTTATGAGCGGCTCCTACTTTCAGTATCGC 86505
QY 423 ProLysProGlnProGlyTyrPheAlaProArgLysIleProThrArgProArgLysMet 442
Db 86506 CCATACGATCGTTCATACAGCTTT----- 86529
QY 443 LeuProLeuLeuIleGlySerAspProLysValGlnGluIleArgArg----- 459
```

```
Db 86530 -----ATCGGTTGATTGATGAGGCTGAAGAGCATGTAAAAACGCCACCAAGAAAG 86577
QY 460 -----HisProSerThrGluTrpLysIleAlaLysGluSer 471
Db 86578 ACATTTCGGCGCAATCCCAATGCACTTCATCTGTCGATCGCTACCGCATGTTCCGGAAGGC 86637
QY 472 ArgValLeuThrAsnLeuLysAsnProSerLeuAlaIle-----Leu 486
Db 86638 AGAAACCTCTTGAACCTGAAATCGAAACCATCTGTCTGCGCGGCTTCTTCGACTCTG 86697
QY 487 PheMetAspAspLysLeuGluAsnThrLeuLysGlyArg-----GlnMetLeuThr 503
Db 86698 GCCAAGAC-----CTGAGCGAGCTATCGCAGCGCGCAACCGAGTTCAGTTTCATACCG 86751
QY 504 AspGluGlnLysGlyArgThrArgValLysThrIleArgAlaLeuProArgLeuPheGly 523
Db 86752 GGTCHAGACGCTGG-----CGGATACATGCG 86778
QY 524 AlaProThrAlaLysAlaGluMetIleAspAlaLysValPheGlnAspIleGluGlu----- 542
Db 86779 CGGAACCTGTATCGTCTGGCAATGAACGACGACCGCGTGTTCGCGCAAGAAAGACCCAC 86838
QY 543 -----ArgProIleProProLeuPhePheGluProLysGlyArg 555
Db 86839 AGAGCGGTGAACGCGTGCCTCCAGCAGGTGATCGACTTGTCCGGCTCAATGGCGGCGAG 86898
QY 556 HisThrArgLeuArgTrpThrGlyAlaAsnGluLysGluIleProGlyLeuGlySerArg 575
Db 86899 AAGATTCACTG-----GCACCTCGCCAGCGCC 86925
QY 576 Phe---IleLeuProSerLeuAspPro---ThrMetProAlaLeuAsnThrAlaPheSer 593
Db 86926 TACACCATTCGGGATGCTCTGATCGTATCAATGTGCTTAACATCATCACCGGCTTACA 86985
QY 594 ThrGlnGlyArgAlaArgAspGluTrpAspThrMetPheLys----- 607
Db 86986 ACGTTTGGT-----AGCCAGATTATGAACCATGTGCAAGCGCGGTTTACACGTTTC 87039
QY 608 -----IleProAsnAsnTrpAsn 613
Db 87040 GAGGCGTCTAGTGTGCCCATTTATCAAAACATGGAAT 87075

RESULT 10
ADCL3546
ID ADCL3546 standard; cDNA; 6126 BP.
XX
AC ADCL3546;
XX
DT
XX
XX 18-DEC-2003 (first entry)
DE Human NOVX protein encoding cDNA sequence, SEQ ID No 25.
KW NOVX; PADD interacting protein; ATPase; H+ Transporting; Lysosomal;
KW FGF 17; Single Pass Transmembrane; Beta-Ketoacyl Synthase; Neurallin 2;
KW Glutamate Receptor Interacting Protein 2; Chr-Methyltransferase;
KW NP25 Variant; GTPase-Activating Protein; ELKS; Sim2; RhoGAP;
KW Phospholipase; Scavenger Receptor Domain Containing Protein;
KW Metallothionein IA; NOGO receptor; FYVE; NOELIN;
KW Cyclin Regulatory Subunit; TetraTico Peptide Repeat Protein;
KW Immunoglobulin Domain Containing Protein; PA Domain Containing Protein;
KW Phenylalanine; Histidine Ammonia-Lyase; Cellular Retinaldehyde-Binding;
KW Glutamine Repeat Containing Protein; TNF Receptor Associated Factor2;
KW Vacuolar Protein Sorting Homologue R-VPS33A;
KW Bola Domain Containing Protein; Neurotrophin Receptor;
KW RAL Guanine Nucleotide Dissociation Stimulator; Armadillo/Beta-Catenin;
KW Metalloprotease; T10 Ser/Thr-rich; Ring finger-like; cytosstatic;
KW gene therapy; vaccine; cancer; gene; ss.
OS Homo sapiens.
XX
FN WO2003004617-A2.
XX
PD 16-JAN-2003.
```



```
Db 1487 -----:::|||||:::|||||
Qy 401 AspArgIleLysTyrValAspValGluLysTyrLeuAlaProLysProMetGlu 420
Db 1520 AGTTTATGACACCAAGATTGATCAGCTGAGACGAGAACTTTCAAGAAAGATCGGAA 1579
Qy 421 PheAsnProLysProGlnProGlyTyrPheAlaProArgLysIlePheProThrArgProArg 440
Db 1580 CTT-----CTTGCTTTACAAACAAAGCTTGAACCCCTCAGCAAT 1618
Qy 441 LysMetLeuProLeuLeuIleGlySerAspProLysValGlnGluIleArgHis 460
Db 1619 CAA-----AATTCAGATTCAAGCAACACATTTGAAGTGTCTCAAGAG 1660
Qy 461 ProSerThrGluTrpLysIleAlaLysGluSerArgValLeuThrAsnLeuLysAsnAsn 480
Db 1661 TCATTACT-----GCCAAGAACAGAGGGTGCATC-----CTTCAGCTGAG 1705
Qy 481 ProSerLeuAlaLeuPheMetAspLys-----LeuGluAsnThrLeuLys 497
Db 1706 GTAGATGCGCTGAGATTACGACTGCAAGAAAGAAATCTTTCTCAATAAACAACAAA 1765
Qy 498 GlyArgGlnMetLeuThrAspGluGlnLysGlyArgThrArgValLysThrIleArgala 517
Db 1766 CAGCTACAGGACCTCACA---GAAGAGAAGGGG-----1795
Qy 518 LeuProArgLeuPheGlyAlaProThrAlaLysAlaGluMetIleAspAlaLysValPhe 537
Db 1796 -----ACCTGCCGCTGAATTCGTGATCAATGAAGATATG 1831
Qy 538 GlnAspIleGluGluArgProIleProLeuPhePheGluProLysGlyArgHisThr 557
Db 1832 TTAGAAGTGAAGAAAGAAATCAATCTTTCAGAAAAGAGATTGAAAACCTGCAAGAA 1891
Qy 558 ArgLeuArgTrpThrGlyAlaAsnGluLysGluIleProGlyLeuGlySerArgPheIle 577
Db 1892 CAACTTAGG-----GATAAAGACAAAGCAACTGACCAACCTGAAAGACAGA-----1936
Qy 578 LeuProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSerThr-----594
Db 1937 GTGAAGTCTTGACAGCGATTCAGTATACAGATACCTGCTGGCAGCCTAGAGAA 1996
Qy 595 -----GlnGlyArgAlaArgAspGlu 601
Db 1997 GCTCTGTGACAGAGAGAGATAATTGAGCGCTTGAAGAACAGCGAGAGAGATGAT 2056
Qy 602 TrpAspThrMetPheLysIleProAsnAsnTrpAsnProGlyAspGluValGlyPheLys 621
Db 2057 CGGGAAGACTAGAGAGATAGATATCTTCGAAAAGAGAACAAAGACCTGAAAGAGAAG 2116
Qy 622 MetAsnSer 624
Db 2117 GTCATGCT 2125

RESULT 11
ABZ11646
ID ABZ11646 standard; cDNA; 6151 BP.
XX
AC ABZ11646;
XX
DT 20-JAN-2003 (first entry)
XX
DE Human polynucleotide SEQ ID NO 528.
XX
KW Human; genome mapping; gene therapy; food supplement; virus; fungus;
KW cell-proliferative disorder; neurodegenerative disease; bacterial;
KW Parkinson's disease; Alzheimer's disease; autoimmune disease;
KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
KW arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
KW haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;
KW antiarthritic; gene; ss.
```

```
XX Homo sapiens.
XX OS WO200270539-A2.
XX PN 12-SEP-2002.
XX PD 05-MAR-2002; 2002WO-US005095.
XX PF 05-MAR-2001; 2001US-00799451.
XX PR (HYSE-) HYSEQ INC.
XX PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao Q, Ren F;
PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
PI Wehrman T, Wang J, Wang D, Drmanac RT;
XX WPI; 2002-759812/82.
XX P-FSDB; ABP69429.
XX New polynucleotides comprising sequences assembled from expressed
XX sequence tags (ESTs), useful for treating cell-proliferative,
XX neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet
XX or coagulation disorders.
XX Claim 1; SEQ ID NO 528; 1012pp + Sequence Listing; English.
XX The invention relates to an isolated polynucleotide (I) comprising a
XX nucleotide sequence selected from any of 948 sequences (ABZ11119-
XX ABZ12066) or their mature protein coding portion, active domain coding
XX protein or complementary sequences. The polynucleotides are useful for
XX identifying expressed genes or for physical mapping of human genome. The
XX encoded polypeptides (ABP68902-ABP6949) are useful as molecular weight
XX markers, as a food supplement, for generating antibodies, in medical
XX imaging, screening and diagnostic assays and for treating cell-
XX proliferative disorders (cancer), neurodegenerative diseases (Parkinson's
XX or Alzheimer's disease), autoimmune diseases (multiple sclerosis,
XX diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,
XX platelet or coagulation disorders, wound, burns, incision, ulcers, liver
XX or lung fibrosis, infections (bacterial, viral, fungal, parasitic),
XX arthritis, etc. Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 6151 BP; 1939 A; 1247 C; 1394 G; 1571 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 0.00177 Length: 6151
Score: 169.50 Matches: 118
Percent Similarity: 36.28% Conservative: 108
Best Local Similarity: 18.94% Mismatches: 248
Query Match: 5.15% Indels: 149
DB: 6 Gaps: 22

US-10-736-868-2 (1-643) x ABZ11646 (1-6151)
Qy 36 LeuValGlyLysSerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGluLeu 55
Db 632 CTGACTGGCTCATCCCATCATCACCACACAGGTCCCTCCCATG---TTGAGCGAGGTA 688
Qy 56 MetGlyValGlnPheValAspAlaLeuIleLysLysGlyGlnMetGluMetAlaLysGly 75
Db 689 AGAGACAGCACAAATGTTAGATCTTCAGGCCCCAGCTGAAAGAACTGCAGAGAGAGATGAC 748
Qy 76 AlaPheLysThrGlnLeuGluValLeuGluLysValHisProAspGlnPheAspLysTyr 95
Db 749 CTCCTCCGGAAGAGCTAGACATCAGGACAGCAATTTGGGATCTTCATCAACAGATT 808
Qy 96 LysLysLeuLysValAspAspLeuAlaAlaAspAlaValMetGlnGlnAlaGluMetAla 115
Db 809 AAGACTTCTGGAGTCTCTGAGTCTTGAAGAGAGAGATCTTGAAGGAAGAGAGAGCGC 868
Qy 116 LysLeuGlnProLysSerGlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIlePro 135
```

```

Db      869  CGGATGCTGTCCTCAGGAGCAGATGAGGTTTCCCATGAAGAAATCAGACCTACAG  928
      ::::
Qy      136  IleGlySerSerIleArgGlyLeuGluAspAlaIleArgThrGlnArgAspMetGluAsn  155
      ::::
Db      929  TTG-----ACAATCCAGGCCCTTCAAGATGAGCTGCGAACCAGAGAGACCTCAACCCAC  982
      ::::
Qy      156  ThrAspProSerGlnGlnIleAlaIleAlaValMetAspLysPheGlnThrGlnIleLeu  175
      ::::
Db      983  CTCCTCCAGCAAGAGAGTGCGCAACCGAGGCG-----GAGCACTTCACCATCGAGCTGACC  1039
      ::::
Qy      176  ProGlyLeuValAlaAsnMetIleAlaGlyIleAsnProPheLysMetProGlnGlnMet  195
      ::::
Db      1040 -----GAGAGAACTTTAGCGCGCTCCAGCCGAG  1069
      ::::
Qy      196  ArgLysAlaGlnAlaAlaProSerSerValPheGlnGlnAlaLeuAlaGlnArgAlaMet  215
      ::::
Db      1070  CATGACAGCGCAGGCTAAGGAGCTGTTCTTTGAGGAAGACATTAGAGGAATGGAGCTG  1129
      ::::
Qy      216  LeuGlyLysAsnAlaProValAlaGlyArgGlyGluGlnArgMetMetMetAsn  235
      ::::
Db      1130 -----AGAATTGAAACGCGAGAAACAACCCCTCAAT  1159
      ::::
Qy      236  ArgValAspGlnArgMetGlnGln-----ArgGluLeuGlnGluAspGluAspAsp  253
      ::::
Db      1160  GCCCGAGATGATCAATTATAAAACTTCTTGAGATGTTGCAAGTAAAGCGCTGCCATCC  1219
      ::::
Qy      254  AspAspLeuGluAspGluAspValProArgArgSerSerAspGlyGluProGln---  272
      ::::
Db      1220  AAAAGCCTGGAGGATGACATAGCGAACGCGCGGATGGCAGAGCGGTGAGTCTCAGGTC  1279
      ::::
Qy      273 -----SerGluAlaGluHisGlnArgArgAsp  281
      ::::
Db      1280  AGCCACTTGAAGTGATTTTAGATCAGAAAGAGAGAAACATACATCTTAGAGAGGAA  1339
      ::::
Qy      282  LeuAlaArgArgLeuLysSerSerProArgLeuLysGluLeuLeuGlnAsnAlaGluVal  301
      ::::
Db      1340  TTGCACCGAAGACCCAACTTCAGCCG-----GAGCCAGCCAGACG  1381
      ::::
Qy      302  GlnSerLeuLeuSerTyArgMetArgAspSerProLeuSerTyArgArgProLeu  321
      ::::
Db      1382  AAGGCTCTCCAGACTGTGCATGAAATGAAGGACACAAAATCGTTCA-----TTG  1432
      ::::
Qy      322  AlaMetAsnAspGluAsp---GluSerAlaPheArgAlaMetGluAlaArgAlaLysLeu  340
      ::::
Db      1433  GAACGNAACATAAGGAGTCTTGAGGATGAGATCCAGATGTAAAGCCAAATGTGTGCTG  1492
      ::::
Qy      341  AspGlnLysSerGlnLeuValLeuGlyLeuHisGlyPheGlyGluSerAspAspGlu  360
      ::::
Db      1493 -----ACACTGAG  1501
      ::::
Qy      361  AspGluGluAspGluAsnLeuIleAspProSerGluAsnSerPheArgArgAlaProLeu  380
      ::::
Db      1502  GACCGCAAGAGAG-----  1516
      ::::
Qy      381  ArgLeuSerSerGlyPheValGluLysLeuLysSerAsnAspGluLeuLysSerAlaLeu  400
      ::::
Db      1517 -----ATCAACAAATTCAGGTTTACAAAGTCATCTCC  1549
      ::::
Qy      401  AspArgIleLysTyArgValAspValGluLysTyArgLeuAlaProLysProMetGlu  420
      ::::
Db      1550  AAGTTTATGAGACCAAGATTGATCAGCTGAGCAGGAACCTTCAAGAAAGAGTCGGAA  1609
      ::::
Qy      421  PheAsnProLysProGlnProGlyTyPheAlaProArgLysIleProThrArgProArg  440
      ::::
Db      1610  CTT-----CTTGCCCTTACAAACAAAGCTTGAACCTCAGCAAT  1648
      ::::
Qy      441  LysMetLeuProLeuLeuIleGlySerAspProLysValGlnGluGluIleArgArgHis  460
      ::::
Db      1649  CAA-----AATTGATGTCAGGCAACACATTTGAAGTGTCTCAAGAG  1690
      ::::
Qy      461  ProSerThrGluTrpLysIleAlaLysGluSerArgValLeuThrAsnLeuLysAsnAsn  480
      ::::

```

```

Db      1691  TCACCTACT-----GCCAAGAACAGAGGGCTGCCATC---CTTCAGACTGAG  1735
      ::::
Qy      481  ProSerLeuAlaAlaLeuPheMetAspAspLys-----LeuGluAsnThrLeuLys  497
      ::::
Db      1736  GTAGATGCGCTGAGATTACGACTGGAGAGAAAAGAAATCTTTCCTCAATAAAAAACAAA  1795
      ::::
Qy      498  GlyArgGlnMetLeuThrAspGluGlnLysGlyArgThrArgValLysThrIleArgAla  517
      ::::
Db      1796  CAGCTACAGACCTCACA---GAAGAGAAGGGG-----  1825
      ::::
Qy      518  LeuProArgLeuPheGlyAlaProThrAlaLysAlaGluMetIleAspAlaLysValPhe  537
      ::::
Db      1826 -----ACACTGGCGGTGAATTCGTGACATGAAGATATG  1861
      ::::
Qy      538  GlnAspIleGluGluArgProLeuPheGluPheGluProLysGlyArgHisThr  557
      ::::
Db      1862  TTAGAAGTGAAGGAGAAATAATCAATGTTCTTCAGAAAAAGATTGAAAACTTGCAAGAA  1921
      ::::
Qy      558  ArgLeuArgTrpThrGlyAlaAsnGluLysGluIleProGlyLeuGlySerArgPheIle  577
      ::::
Db      1922  CAACTTAGG-----GATAAAGCAAGCAACTGACCAACCTGAAGACAGA-----  1966
      ::::
Qy      578  LeuProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSerThr-----  594
      ::::
Db      1967  GTGAAGTCTCTGCAGACGGGATTCACGTAATACAGATACTACTGCTGCGCAGCTAGAGGAA  2026
      ::::
Qy      595 -----GlnGlyArgAlaArgAspGlu  601
      ::::
Db      2027  GCTCTGTCAGAGAGAGAGAAATAATTGAGCGCTTGAAAGAACAGCAGAGAGATGAT  2086
      ::::
Qy      602  TrpAspThrMetPheLysIleProAsnAsnTrpAsnProGlyAspGluValGlyPheLys  621
      ::::
Db      2087  CGGAAAGACTTGAAGAGATAGATCCTTCCGAAAAGAGAACAAAGACCTGAAAGAGAAC  2146
      ::::
Qy      622  MetAsnSer 624
      ::::
Db      2147  GTCAATGCT 2155
      ::::
RESULT 12
ADM44164
ID      ADM44164 standard; cDNA; 6151 BP.
XX
AC      ADM44164;
XX
DT      03-JUN-2004 (first entry)
XX
DE      Novel human arginine-rich protein cDNA #528.
XX
KW      ss; gene; human; arginine-rich protein; cancer; inflammation;
KW      genetic disorder.
XX
OS      Homo sapiens.
XX
PN      US2004053250-A1.
XX
PD      18-MAR-2004.
XX
PF      21-NOV-2002; 2002US-00302172.
XX
PR      05-MAR-2001; 2001US-00799451.
PR      05-MAR-2002; 2002WO-US005095.
PR      20-AUG-2002; 2002US-00225251.
XX
PR      (TANG/) TANG Y T.
PA      PA (XUEA/) XUE A.
PA      (DRMA/) DRMANAC R T.
XX
PI      Tang YT, Xue A, Drmanac RT;
XX
XX      WPI; 2004-238579/22.
XX
PT      New isolated arginine-rich protein-like polynucleotides and polypeptides,
PT      useful for diagnosing and/or treating conditions associated with aberrant

```

PT activity of the arginine-rich polypeptides, such as cancer and
 XX inflammation.

PS Disclosure; SEQ ID NO 528; 51pp; English.

XX The invention relates to an isolated polynucleotide. The methods and
 CC compositions of the present invention are useful for the diagnosis and/or
 CC treatment of diseases or conditions associated with aberrant expression
 CC or activity of the arginine-rich protein-like polypeptides, such as
 CC cancer and inflammation. They can also be used in forensics, gene
 CC mapping, identification of mutations responsible for genetic disorders,
 CC and in assessing biodiversity. The present sequence represents a novel
 XX human arginine-rich protein cDNA.

SQ Sequence 6151 BP; 1939 A; 1247 C; 1394 G; 1571 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0.00177 Length: 6151
 Score: 169.50 Matches: 118
 Percent Similarity: 36.28% Conservative: 108
 Best Local Similarity: 18.94% Mismatches: 248
 Query Match: 5.15% Indels: 149
 DB: 12 Gaps: 22

US-10-736-868-2 (1-643) x ADM44164 (1-6151)

QY 36 LeuValGlyLysSerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGluLeu 55
 DB |||||
 QY 632 CTGACTGGCTATCCATCCATCATCACCACCAGGTCCTCCATG---TTGAGGCGAGTA 688
 DB |||||
 QY 56 MetGlyValGlnPheValAspAlaLeuLeuLysGlyGlnMetGluMetAlaLysGly 75
 DB |||||
 QY 689 AGAGACAGCAATCTTAGATCTTCAGGCCAGCTGAAAGAACTGCAGAGAGAGATGAC 748
 QY 76 AlaPheLysThrGlnLeuGluValLeuGluLysValHisProAspGlnPheAspLysTyr 95
 DB |||||
 QY 749 CTCCTCCGGAAGAGCTAGATAGACAGACAGCAAAATGGGATCTTCATGAACAGATAT 808
 QY 96 LysLysLeuLysValAspLeuAlaAlaAspAlaValMetGlnGlnAlaGluMetAla 115
 DB |||||
 QY 809 AAGACTTCTGGAGTCCTGAGCTTAAGAGAGAGAGAGCTTTGAGGAAGAGAGAGCGCG 868
 QY 116 LysLeuGlnProLysSerGlyAsnAlaPheLeuAspMetLeuAsnGlyAsnGlyPro 135
 DB |||||
 QY 869 CGGATGTCTGCTCTCAAGGAGCAGATGAGGTTTCCCATGAAGAAATCAGCACCTACAG 928
 QY 136 IleGlySerSerIleArgGlyLeuGluAspAlaIleArgThrGlnArgAspMetGluAsn 155
 DB |||||
 QY 929 TTG-----ACAATCCAGGCCCTTCAGATGAGCTGCGAACCCAGAGAGACCTCAACCC 982
 QY 156 ThrAspProSerGluGlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeu 175
 DB |||||
 QY 983 CTCCTCCAGCAGAGATGGCAACCGAGGCG---GAGCACTTCACCATCGCTGACC 1039
 QY 176 ProGlyLeuValAlaAsnMetIleAlaGlyLysAsnProPheLysMetProGlnGlnMet 195
 DB |||||
 QY 1040 -----GAGGAGAACTTTAGCGCGCTCCAAAGCCGAG 1069
 QY 196 ArgLysAlaGlnAlaAlaProSerSerValPheGlnGlnAlaLeuAlaGlnArgAlaMet 215
 DB |||||
 QY 1070 CATGACAGCGCGGTGAAGAGCTGTCTCTTGAAGAGACATAGAGAAATGGAGCTG 1129
 QY 216 LeuGlyLysAsnAlaProValalaglyGlyArgGlyGluGlnArgMetMetMetAsn 235
 DB |||||
 QY 1130 -----AGAATTGAACGCGAACAACAACCCCTCAAT 1159
 QY 236 ArgValAspGlnArgMetGlnGln-----ArgGluLeuGlnGluAspGluAspAsp 253
 DB |||||
 QY 1160 GCCGAGATGATCAATTAATAAACTTCTTGATGATGTTGCAAGTAAGGCTTGCCATCC 1219
 QY 254 AspAspLeuGluAspGluAspValProArgArgArgSerAspGlyGluProGln--- 272
 DB |||||
 QY 1220 AAAAGCCTGGAGGATGACATGAGCGAACCGCGGGGATGGCAGAGGCTGAGTCTCAGGTC 1279

QY 273 -----SerGluAlaGluHisGlnArgAsp 281
 DB |||||
 QY 1280 AGCCACTTGGAGTCAATTTTAGATCAGAAAGAGGAAGAAACATACATCTTAGAGAGNA 1339
 QY 282 LeuAlaArgArgLeuLysSerSerProArgLeuLysGluLeuLeuGlnAsnAlaGluVal 301
 DB |||||
 QY 1340 TTGCACCGAAGAACCACTTCCAGCG-----GAGCCAGCCCAAGACG 1381
 QY 302 GlnSerLeuLeuSerTyrGlnArgMetArgAspSerProLeuSerLysArgArgProLeu 321
 DB |||||
 QY 1392 AAGGCTCTCCAGACTGTCTCATGAATGAAGACACAAAAATCGCTTCA-----TTG 1432
 QY 322 AlaMetAsnAspGluAsp---GluSerAlaPheArgAlaMetGluAlaArgAlaLysLeu 340
 DB |||||
 QY 1433 GAACAAACATAAGGATCTTGAGATGAGATCCAGATGTTTAAAGCCATCGTGTGTG 1492
 QY 341 AspGlnLysSerGlnLeuValLeuGlyLeuHisGlyPheGlyGluSerAspAspGlu 360
 DB |||||
 QY 1493 -----AACACTGAG 1501
 QY 361 AspGluGluAspGluAsnLeuLeuAspProSerGluAsnSerPheArgArgAlaProLeu 380
 DB |||||
 QY 1502 GACCGCGAAGAGAG----- 1516
 QY 381 ArgLeuSerSerGlyPheValGluLysLeuLysSerAsnAspGluLeuLysSerAlaLeu 400
 DB |||||
 QY 1517 -----ATCAACAATTTAGGTTTACAAAGTCACTCC 1549
 QY 401 AspArgIleLysTyrArgValAspValGluLysTyrLeuAlaProLysProMetGlu 420
 DB |||||
 QY 1550 AAGTTTATGAAGCAAGATTGATCAGCTGAAGCAGGAACCTTCAAGAAAGAGTCGAA 1609
 QY 421 PheAsnProLysProGlnProGlyTyrPheAlaProArgLysIleProThrArgProArg 440
 DB |||||
 QY 1610 CTT-----CTTGCCCTTCAACAACAAAGCTTGAACCCCTCAGCAAT 1648
 QY 441 LysMetLeuProLeuLeuIleGlySerAspProLysValGlnGluGluIleArgArgHis 460
 DB |||||
 QY 1649 CAA-----AATTCAGATTGCAAGCAACACATTTGAAGTCTCAAGAG 1690
 QY 461 ProSerThrGluTrpLysIleAlaLysGluSerArgValLeuThrAsnLeuLysAsnAsn 480
 DB |||||
 QY 1691 TCACCTACT-----GCCAAGAACAGAGGCTGCCATC---CTTCAGACTGAG 1735
 QY 481 ProSerLeuAlaAlaLeuPheMetAspLys-----LeuGluAsnThrLeuLys 497
 DB |||||
 QY 1736 GTAGATCGCTGAGATTACGACTGGAAGAAAGAAATCTTTCTCAATAAAAAACAATA 1795
 QY 498 GlyArgGlnMetLeuThrAspGluGlnLysGlyArgThrArgValLysThrIleArgAla 517
 DB |||||
 QY 1796 CAGTACAGGACCTCACA---GAAGAGAGAGGG----- 1825
 QY 518 LeuProArgLeuPheGlyAlaProThrAlaLysAlaGluMetIleAspAlaLysValPhe 537
 DB |||||
 QY 1826 -----ACACTGCCGCTGAAATTCGTGACATGAAAGATATG 1861
 QY 538 GlnAspIleGluArgProIleProLeuPhePheGluProLysGlyArgHisThr 557
 DB |||||
 QY 1862 TTAGAAGTGAAGGAAGAAATAATCAATGTTCTTCAGAAAAAGATTGAACCTTCAAGAA 1921
 QY 558 ArgLeuArgTrpThrGlyAlaAsnGluLysGluIleProGlyLeuGlySerArgPheIle 577
 DB |||||
 QY 1922 CAACCTTAG-----GATAAGACAGCAACTGACCAACCTTGAAGACAGA----- 1966
 QY 578 LeuProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSerThr----- 594
 DB |||||
 QY 1967 GTGAAGTCTTTCAGACGGATTCCAGTAATACAGATACTGCTGCTGGCGCGCTAGAGAA 2026
 QY 595 -----GlnGlyArgAlaArgAspGlu 601
 DB |||||
 QY 2027 GCTCTGTCAGAGAGGAGAGAATAATTGAGCGCTTGAAGAAAGAACGAGGAAGAGATGAT 2086

```
QY 602 TrpAspThrMetPheLysIleProAsnAsnTrpAsnProGlyAspGluValGlyPheLys 621
Db 2087 CGGGAAGACTGAGAGATAGATCTCTCCGAAAAGAGAACCAAGACCTGAAAGAGAG 2146
QY 622 MetAsnSer 624
Db 2147 GTCATGCT 2155

RESULT 13
ID ABL02411 standard; cDNA; 5439 BP.
AC ABL02411;
XX
XX 26-MAR-2002 (first entry)
DT
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 1715.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX
XX 11-JUN-2000; 2000US-00614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX P-PSDB; ABB9308.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.
XX
XX Claim 1; SEQ ID NO 1715; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
XX sequences (ABU1840-ABU16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 5439 BP; 1803 A; 1228 C; 1463 G; 945 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 0.00201 Length: 5439
Score: 168.00 Matches: 124
Percent Similarity: 36.36% Conservative: 100
Best Local Similarity: 20.13% Mismatches: 230
Query Match: 5.10% Indels: 162
DB: 4 Gaps: 29

US-10-736-868-2 (1-643) x ABL02411 (1-5439)

QY 65 IleLysLys-----GlyGlnMetGluMetAlaLysGlyAlaPheLysThrGlnLeuGlu 82
Db 1102 CTTAAGAAAATGAGCAGTGAATAATGAGCTGGAAATGGCGGATGAAGGAAGACGAGAA 1161
QY 83 ValLeuGluLysValHisProAspGlnPheAspLysTyrLysLysLeuLysValAspAsp 102
```

```
Db 1162 ATGCAGAA-----CCTCCAAAAGGCAAGAAAAATAAACAAACGAATGGACAGT 1212
QY 103 LeuAlaAlaAspAlaValMetGlnGlnAlaGluMetAlaLysLeuGlnProLysSerGly 122
Db 1213 GAAAGCGAAGATGAAATC---CCGAAAACCGAAGCGAGAAAATAACCAAGTTCCACCCAAA 1269
QY 123 AsnAlaPheIleAspMetLeuAsnGlyAsnGly----- 133
Db 1270 AACAAACTGAAAGGATTGGTGGACTCCGAATCTGAACCGGAGAAACAACTCTGAGGAAGTT 1329
QY 134 IleProIleGlySerSerIleArgGlyLeuGluAspAlaIleArgThrGlnArgAspMet 153
Db 1330 AGTCCCGTCAAGNACAAACTTAAGAGATTGGTGACTCG---GAATCGGAACACAGAGCTA 1386
QY 154 GluAsnThrAspProSerGluGlnIleAlaLysAlaValMetAspLysPhe----- 170
Db 1387 GACAATCCCGAGGAATCTGCAGGAGAACAGGAAGCCCAATGGAAGCGCTCTCTCCAGA 1446
QY 171 -----GlnThrGlnIleLeuProGlyLeuValAlaAsnMetIleAlaGlyLys 186
Db 1447 GAGAAACCTAAGAAAGCCAAAGTTGTCGGAATCAGCTAAGAAAGACACTGGAGGGGATG 1506
QY 187 AsnProPheLysMetProGlnGlnMet-----ArgLysAla----- 198
Db 1507 CAGGCCATCCAAAGTGAACAGCAGCGCTTTCACCGGAGGCACACATCAACGTGCCGTAT 1566
QY 199 ---GlnAlaAlaProSerSerValPheGlnGlnAlaLeuAlaGlnArgAlaMetLeuGly 217
Db 1567 CATCAGCCGAAACCAAGGACACTG---AAGGAGTTTCTTAGCCGAGCAACCATC----- 1617
QY 218 LysAsnAlaPro-----ValAlaGlyGlyArgGlyGluGlnArgMetMet 233
Db 1618 ---AACGCTCCACTGGCCACGGCTCTGGCTGGAGGAAGTCCATGATGCCAGCGAGGCCA 1674
QY 234 MetAsnArgValAspGlnArgMetGlnArgGluLeuGlnGluAspGluAspAsp 253
Db 1675 AGGAAGTCGTGGACTGCGGATGACCAGGAGGAAATGGAGGCTTATGCAAACTATG 1734
QY 254 AspAspLeuGluAspGluAspValProArgArgSer-----SerAspGlyGlu 270
Db 1735 GAGGATCGTCCCAAGAAAGCTACCGAGTTCTTTAAGTCTGAGTCTGAGTCAGATGAGGAG 1794
QY 271 ProGlnSerGluAlaGluHisGlnArgArgAspLeuAlaArgArgLeuLysSerSerPro 290
Db 1795 GATGATTCGAGAAACGAAGAA-----CCTATGGAATTAAAGCAATCCA 1839
QY 291 ArgLeuLysGluLeuLeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrGlnArgMet 310
Db 1840 GGTGTTATGAT-----GAGATG 1857
QY 311 ArgAspSerProLeuSerLysArgArgProLeuAlaMetAsnAspGluAspGluSerAla 330
Db 1858 CTGATAAATCCAAAACCTCCAGAACAGCCC-----AAGAACGACGACAGAG----- 1902
QY 331 PheArgAlaMetGluAlaArgAlaLysLeuAspGlnLysSerGlnLeuValLeuGlyLeu 350
Db 1903 -----GATGACCCAGCCAGCAGTTATTTAGGTA 1932
QY 351 HisGlyPheGlyGluSerAspAspGluAspGluGluAspGluAsnLeuLe----- 368
Db 1933 CCTGTCTCTGGTCAAGACAAACACAGATGATACCGTGGTGATGAAGCAATGGTTGCAGAT 1992
QY 369 -----AspProSerGluAsnSerPheArgArgAlaProLeuArgLeuSer 383
Db 1993 TCTATTACTGAAGAAGAACCCATTGGCTTCCACCTCTACTGCAGCTGCCCTGAAATAGCA 2052
QY 384 SerGlyPheValGluLysLeuLysSerAsnAspGluLeuLysSerAlaLeuAspArgIle 403
Db 2053 GATAAATTTTGAATTCAGCAAGAGACTAATGCGAAATCCCCAGCAAAAGTCTGTCTGGTT 2112
QY 404 LysTyrArgValAspAspValGluLysTyrLeuAlaProLys-----Promet 419
|||
```


Qy	199	GlnAlaAProSerValPheGlnGlnAlaLeuAlaClnArgAlaMetLeuGlyLys	218
Db	5011	CTCGAAGCCGGAAGACGAGCTCCAGGTCTGCCTCGAGGAGCTGAG	
Qy	219	AsnAlaProValIalaglyArgGlyGluGluGlnArgMetMet	233
Db	5059	-----GCCGCTCTTGACGAGGAGAACCAAGGTGCTCCGCGCTCAGCTTGAG	5106
Qy	234	MetAsnArgValAspGlnArgMetGlnGlnArgGluLeuGlnGlnGluAspGluAspAsp	253
Db	5107	CTGTCCAGTGGCCAGAGATCGACCGCCGC---ATCCAGAGAAGGAGGAGGATTTC	5163
Qy	254	AspAspLeuGluAspGluAspValProArgArgSerSerAspGlyGluProGlnSer	273
Db	5164	GAGAACACCCGCAAGAACACCACGCTGCCCTCGACTCCATGCGAGGCTTCCCTCGAAGCC	5223
Qy	274	GluAlaGluHisGlnArgArgAspLeuAlaArgArgLeuLysSerSerProArgLeuLys	293
Db	5224	GAGGCCAAGGGCAAGCTGAGGCCCTCGCATGAAGAAGAAGCTGAGGGTGACATCAAC	5283
Qy	294	GluLeu-----LeuGlnAsnAlaGluValGln-----	302
Db	5284	GAGCTTTGAGATTGCTCTGGATCACGCCAAACAAGGCTTAACGCCGAGGCCCAAGAGAATC	5343
Qy	303	-----SerLeuLeuSerTyrGlnArgMet	310
Db	5344	AACGGTTACGAGCAGCTGAAGGACATCCAGACTGCGCCTCGAGGAGGACGCGGCC	5403
Qy	311	ArgAspSerPro-----LeuSerLysArgArgProLeuAlaMetAsnAsp	325

3404	CGCGAGATGCGCGCGAGACGTGGGGATGTCGAGGCGTGCTGCGAAACGCGCTTCCAGAAC	3406	
Qy	326	GluAspGluSerAlaPheArgAlaMetGlu-----	335
Db	5464	GAACCTGGAGGAGTCTCGCATCTGCTGGAAACAGGCGCGTGGCGTGCAGGCGCGAA	5523
Qy	335	-----	335
Db	5524	CAGGAGCTGGCGGATGCCACGAGCAGCTGAACGAAAGTGTGCGCCGAGAACGCTCCATC	5583
Qy	336	-----AlaArgAlaLysLeuAspGlnLysSerGlnLeuValLeuGlyLeuHisGlyPhe	353
Db	5584	TCCGCTGCCAAGAGGAGCTGGAGTCCGAGCTCCAGACC-----CTGCAC-----	5628
Qy	354	GlyGluSerAspAspAspGluAspGluGluAspGluAsnLeuLeuAspProSerGluAsn	373
Db	5629	-----TCCGACCTGCGCAA-----CTCTGAACGAAGCAAGAAC	5664
Qy	374	SerPheArgArgAlaProLeuArgLeuSerSerGlyPheValGlnLysLeuLysSerAsn	393
Db	5665	TCCGAGGAAGGCC-----AAGAGGCTATGCTCGATCCGCCCGCTGGGCC	5712
Qy	394	AspGluLeuLysSerAlaLeuAsp-----	401
Db	5713	GATGAGCTCGCGCTGAGCAGGATCATGCCACAGCCAGGAGAAATTGAGGAGGCCCTC	5772
Qy	402	-----ArgIleLysTyrArgValAspValGlnLysTyrLeuAlaPro	416
Db	5773	GAGCAGCAGATCAAGAGAGCTGCAGGTCCGCTCTGGACGAGGCTCAGGCCAACGCCCTC---	5829
Qy	417	LysProMetGluPheAsnProLysProGlnProGlyTyrPheAlaProArgLysIlePro	436
Db	5830	-----AAGGAGGCGCAAGAGGCCATTCAAGAGCTTGAG	5862
Qy	437	ThrArgProArgLysMetLeuProLeuIleGlySerAspProLysValGlnGluGlu	456
Db	5863	CAGCGCGTCCGCGAGCTC-----GAGAACCGAGCTGGATGCTGAG	5901
Qy	457	IleArgArgHisProSerThrGluTrpLysIleAlaLysGluSerArgValLeuThrAsn	476
Db	5902	CAGAGGAGGACGCCCGATGCCCAGAGAAACCTTCGCGAAGTCCGAGGCTCGCGTCAAGGAG	5961

```

Db      5962 CTGAGCTTCAGTCCGAGGAGGACCGCAAGAACCCAGAGCGCATGCGGATCTGGTC--- 6018
Qy      489 AspAspLysLeuGluAenThrLeuLysGlyArgGlnMetLeuThrAspGluGlnLysGly 508
Db      6019 ---GACAAAGCTGCAACACAGAGATCAAGACATACAGAGAGCGCAGATCGAGGAG----- 6066
Qy      509 ArgThrArgVallysThrIleArgAlaLeuProArgLeuPheGlyAlaProThrAlaLys 528
Db      6067 -----GCTGAGGAAATCGCGCCCTC---AACTTGGCCAAATTCGCAAGGCTCAG 6114
Qy      529 AlaGluMetIleAspAlaLysValPheGlnAspIleGluGluArgProIleProProLeu 548
Db      6115 CAGGAGCTTGAGGAGCGCGAGGAGCGCGCGATCTGGCCGAGCAGGCCATCAGCAAA--- 6171
Qy      549 PhePheGluProLysGlyArg 555
Db      6172 ---TTCCGCGCCCAAGGACGT 6189

```

Search completed: July 15, 2005, 16:01:48
Job time : 1204 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 18, 2005, 12:25:31 ; Search time 305 Seconds
(without alignments)
3449.593 Million cell updates/sec

Title: US-10-736-868-2
Perfect score: 3291
Sequence: 1 MILEFLFLLLGFCIAPLSA.....SKTRFVGNGAFDMPALGL 643

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO spool/US10736868/runat_14072005_105510_2303/app_query.fasta_1.839
-DB=Issued_Patents_NA -QMT=fastcap -SUFFIX=rn1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10736868 @CGN 1.1.69 @runat_14072005_105510_2303 -NCPU=6 -ICPU=3
-NO_MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=110 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTCUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	171	5.2	100990	4	US-09-409-800B-2
2	169.5	5.2	6151	4	US-09-799-451-528
3	162.5	4.9	5883	4	US-09-949-016-5001
4	159	4.8	2680	2	US-08-533-306A-5
5	159	4.8	2680	2	US-08-742-923A-5
6	159	4.8	2887	2	US-08-533-306A-3
7	159	4.8	2887	2	US-08-742-923A-3
8	159	4.8	4935	4	US-09-949-016-1019
9	159	4.8	4935	4	US-09-949-016-5054
10	159	4.8	5574	4	US-09-917-254-40
11	159	4.8	6861	4	US-09-949-016-1240
12	159	4.8	6861	4	US-09-949-016-1241

13	159	4.8	6861	4	US-09-949-016-1242
14	158.5	4.8	19269	4	US-09-902-540-1175
15	157	4.8	6175	4	US-08-875-435B-1
16	157	4.8	6856	4	US-09-566-921-42
17	156	4.7	3807	4	US-09-645-456A-8
18	156	4.7	3807	4	US-09-425-324A-8
19	156	4.7	3807	4	US-09-645-791-8
20	155.5	4.7	3972	4	US-09-645-456A-6
21	155.5	4.7	3972	4	US-09-425-324A-6
22	155.5	4.7	3972	4	US-09-645-791-6
23	153.5	4.7	6396	4	US-09-949-016-3344
24	153	4.6	9551	1	US-08-056-200-93
25	153	4.6	9551	2	US-08-800-644-93
26	152.5	4.6	1761	4	US-09-252-991A-16401
27	152.5	4.6	3489	4	US-09-252-991A-16193
28	152.5	4.6	4055	4	US-09-688-188B-10
29	152.5	4.6	4055	4	US-09-291-417D-10
30	152	4.6	3831	4	US-09-645-456A-5
31	152	4.6	3831	4	US-09-425-324A-5
32	152	4.6	3831	4	US-09-645-791-5
33	151.5	4.6	3996	4	US-09-645-456A-2
34	151.5	4.6	3996	4	US-09-425-324A-2
35	151.5	4.6	3996	4	US-09-645-791-2
36	151.5	4.6	6644	4	US-08-875-435B-5
37	150.5	4.6	3228	4	US-09-252-991A-1816
38	150.5	4.6	3546	4	US-09-252-991A-2143
39	150.5	4.6	4209	4	US-09-248-796A-400
40	150.5	4.6	7596	4	US-09-023-655-1463
41	149	4.5	2130	4	US-09-248-796A-4942
42	148.5	4.5	2094	4	US-09-248-796A-1868
43	146.5	4.5	3894	4	US-09-645-456A-7
44	146.5	4.5	3894	4	US-09-425-324A-7
45	146.5	4.5	3894	4	US-09-645-791-7

ALIGNMENTS

RESULT 1
US-09-409-800B-2
; Sequence 2, Application US/09409800B
; Patent No. 6706322
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; APPLICANT: Burland, Valerie
; APPLICANT: Rose, Debra J.
; APPLICANT: Mayhew, George F.
; APPLICANT: Perna, Nicole
; APPLICANT: Perry, Robert D.
; APPLICANT: Straley, Susan C.
; APPLICANT: Fetherston, Jacqueline D.
; APPLICANT: Lindler, Luther E.
; APPLICANT: Plano, Gregory V.
; TITLE OF INVENTION: Plasmid DNA From Yersinia Pestis
; FILE REFERENCE: 960296.95939
; CURRENT APPLICATION NUMBER: US/09/409,800B
; CURRENT FILING DATE: 1999-09-30
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 100990
; TYPE: DNA
; ORGANISM: Yersinia pestis
US-09-409-800B-2

Alignment Scores:
Pred. NO.: 2.82e-05
Score: 171.00
Matches: 142
Percent Similarity: 33.99%
Conservative: 100
Best Local Similarity: 19.94%
Mismatches: 260
Query Match: 5.20%
Indels: 210
DB: 33
Gaps: 33
US-10-736-868-2 (1-643) x US-09-409-800B-2 (1-100990)

Qy 17 ProLeuSerAlaGlnSerProSer---ThrSerAspAlaProGlyAlaLeuLeuSerSer 35
Db 85225 CCAAGTGGTGTAAACATCCCATCTATATACCGGACGACGCATCAACCAACCTCATGAACGCA 85284
Qy 36 LeuValGly----- 38
Db 85285 GTACGTGGTCTTCGACCATGAAGTAGCCCATATCTCTTTACTGATCCGAAAGTCGCG 85344
Qy 39 ---LysSerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGluLeu 55
Db 85345 ATGAAGATGCGCGAGAGGAAAGGCTCTCTACCGGGCTTTGGAAACGCGCTGGAA--- 85401
Qy 56 MetGlyValGlnPheValAspAlaLeuLeuLysLysGlyGlnMetGluMetAlaLysGly 75
Db 85402 ---GACGTGTTTATTTGAACAGAGATGGCAGGTATTTCAACGGA 85443
Qy 76 AlaPheLysThrGlnLeuGluValLeuGluLysValHisProAspGlnPheAspLysTyr 95
Db 85444 ACCGTCGTAAATTGTCACACAGAACCTGGTG-----ATCGAACAAATAC 85491
Qy 96 LysLysLeuLysValAspAlaLeuAlaAspAlaValMetGlnGlnAlaGluMet--- 114
Db 85492 TTCAAGGGCAAGTGTACAGAGCGGTTTCAATCTGCCACGCAACACGCGTGAATTGTTTC 85551
Qy 115 ---AlaLysLeuGlnProLysSerGlyAsnAla---PheIleAsp 127
Db 85552 CTGAATAATCTTCCTTTGCCCGTGTCCGCGCTGGGATGGCCAAAGCCCTTTTCATCGAC 85611
Qy 128 MetLeuAsnGlyAsnGlyIleProIleGlySerSerIle--- 142
Db 85612 TTTATGGAAACACTGGCACCTCATCGATAAGCCTGTACGCTTGTCTCAAGAGCATGGT 85671
Qy 143 LeuGluAspAlaIleArgThrGlnArgAspMetGluAsnThrAspProSerGluGlnIle 162
Db 85672 ATCGACGTGGCGGTT---CGCAATATGTGCAACACAGAGGACTCGGTAAGGTC 85722
Qy 163 AlaLysAlaValMetAspLysPheGlnThrGlnIleLeuProGlyLeuValAlaAsnMet 182
Db 85723 GCAGCTCTATCGCCCAATCATCGCAGAT--- 85752
Qy 183 IleAlaGlyLysAsnProPheLysMetProGlnMetArgLysAlaGlnAlaAlaPro 202
Db 85753 ATGAAGACAGCGCAGAGGCAATTAACCT---GAGCTTAATCATCTGCCAAAGCCG 85809
Qy 203 SerSerValPheGlnGlnAlaLeuAlaGlnArgAlaMetLeuGlyLysAsnAlaProVal 222
Db 85810 TCTAAGAGCAAGACGAGTCC-----GAGGAAACACACAGAG 85845
Qy 223 AlaGlyGlyArgGlyGluGluGlnArgMetMetAsnArgValAspGlnArgMetGln 242
Db 85846 TCAGGCGCAGATCCGTCCTCATAGTAATCTGCAACAAAGCGCAACCAAGCGAAGCAGAC 85905
Qy 243 GlnArgGluLeuGlnGluAsp---GluAspAspAspLeuGluAspGluAspVal 261
Db 85906 GACAAGAGAGCAAGAAAGATGATCCCTCAGAGAGAGAAATCTGGGGATTCGTATTG 85965
Qy 262 ProArgArgArgSerAspGlyGluProGlnSerGluAlaGluHisGlnArgArgAsp 281
Db 85966 CCT-----GATTCATTAGATAAGGACTTACCTATACATGAT----- 86001
Qy 282 LeuAlaArgArgLeuLysSerSerProArgLeuLysGluLeuLeuGlnAsnAlaGluVal 301
Db 86002 -----AAAGAAATTTAGTGATACAGAAAGTAAA 86028
Qy 302 GlnSerLeuLeuSerTyrGlnArgMetArgAspSerProLeuSerLysArgArgProLeu 321
Db 86029 TATACAGACGAGCGGAAGATGAGTCAGGAGACACCCCAAGATCCGATGTCGCGCATG 86088
Qy 322 AlaMetAsnAsp---GluAspGluSerAlaPheArgAlaMetGluAla----- 336
Db 86089 GAATCAGGTGATCTGACGACGAGAGGTGGTAGTGATGGCACTGGCGCTCCCAACGCGCTGT 86148

RESULT 2

Qy 337 -----ArgAlaLysLeuAspGlnLysSerGlnLeuValLeuGlyLeuHisGlyPhe 353
Db 86149 GATGGCATTCGGCAAGACCCGATGATCTCGATGTTATGGCTCTGCGCGCTGGCGAT 86208
Qy 354 GlyGluSerAsp-----AspAspGluAsp 361
Db 86209 GGGGATAGTGATAACCGCAAGACTCCGATGCGGTCGTGAGGAGTCCGAAAGGAGAAGAC 86268
Qy 362 GluGluAspGluAsnLeuIleAspProSerGluAsnSerPheArg----- 376
Db 86269 GAAGCGGAAGAGGACGTCCCGCATCACTACGCGCAAGGCAAGAAATGAGGATGCC 86328
Qy 377 ---ArgAlaProLeuArgLeuSerSerGlyPheVal-----GluLysLeuLysSer 392
Db 86329 GCGGAAGCGCTCGAAGACAGTGAAGTCTGCTGCGCGGATGAATGACTCTG 86388
Qy 393 AsnAspGluLeuLysSerAlaLeuAspArgIleLysTyrArgValAspAspVal----- 410
Db 86389 GAGGACGCACTCAAG---GGCTCGACGAGATGGAAGAAGAACACAGCAAAATGACCGAA 86445
Qy 411 -----GluLysTyrLeuAlaProLysProMetGluPheAsn 422
Db 86446 GACGCACTGTCCGCCACCATCAGCAAGAGCTTATAGCGCCTCACTTTCTGAGTATCGC 86505
Qy 423 ProLysProGlnProGlyTyrPheAlaProArgLysIleProThrArgProArgLysMet 442
Db 86506 CCATACGATCGTTTCATACGACTTT----- 86529
Qy 443 LeuProLeuLeuIleGlySerAspProLysValGlnGluIleArgArg----- 459
Db 86530 -----ATCGGTTGATTGATGAGCTGGAAGAGCATGTAAACCGCACCAGAAAG 86577
Qy 460 -----HisProSerThrGluTrpLysIleAlaLysGluSer 471
Db 86578 ACATTCGGCGCAATCCCAATGCACTCACCTGTGATCGCTACCGCATGGTTCCGGAAGGC 86637
Qy 472 ArgValLeuThrAsnLeuLysAsnAsnProSerLeuAlaAla-----Leu 486
Db 86638 AGAAACTCTTGAACCTGAAATCGAAACATCTGTCTGCGCGCGTTCCTCGACTCTG 86697
Qy 487 PheMetAspAspLysLeuGluAsnThrLeuLysGlyArg-----GlnMetLeuThr 503
Db 86698 GCCAAAGAC-----CTGGAGCGAGTATCGCCAGCGCAACCGAGTTCAGTTTCATACCG 86751
Qy 504 AspGluGlnLysGlyArgThrArgValLysThrIleArgAlaLeuProArgLeuPheGly 523
Db 86752 GGTCAAGACGTTGG-----CGGATACATGGC 86778
Qy 524 AlaProThrAlaLysAlaGluMetIleAspAlaLysValPheGlnAspIleGluGlu--- 542
Db 86779 CGAACTGTATCGTCTGCAATGAACGACCGCGGTGTTCCGCAAGAAAGAACCCAC 86838
Qy 543 -----ArgProIleProProLeuPhePheGluProLysGlyArg 555
Db 86839 AGAGCGGTGAACGCGTCCGAGAGGTGATCGACTTGTGCGGCTCAATGGGCGGCGAG 86898
Qy 556 HisThrArgLeuArgTrpThrGlyAlaAsnGluLysGluIleProGlyLeuGlySerArg 575
Db 86899 AAGATTCAACTG-----GCATCTCGCAGCGCC 86925
Qy 576 Phe---IleLeuProSerLeuAspPro---ThrMetProAlaLeuAsnThrAlaPheSer 593
Db 86926 TACCACTTGGGATGCTCTGGATCGTATCAATGCTTAACATCATCACCCTTTTACA 86985
Qy 594 ThrGlnGlyArgAlaArgAspGluTrpAspThrMetPheLys----- 607
Db 85986 ACCTTTGGT-----AGCCAGATTTATGAACCATGTGGAAGCGCGGTTTACACGCTTC 87039
Qy 608 -----IleProAsnAsnTrpAsn 613
Db 87040 GAGCGCTCATGTGCTCCCATTTATCAAAAACCTGGAAT 87075

US-09-799-451-528

; Sequence 528, Application US/09799451

; Patent No. 6783969

; GENERAL INFORMATION:

; APPLICANT: Tang, Y. Tom

; APPLICANT: Zhou, Ping

; APPLICANT: Goodrich, Kyle

; APPLICANT: Asundi, Vinod

; APPLICANT: Ren, Feiyang

; APPLICANT: Zhang, Jie

; APPLICANT: Xue, Aidong J.

; APPLICANT: Zhao, Qing A.

; APPLICANT: Wang, Jian-Rui

; APPLICANT: Ma, Yunqing

; APPLICANT: Yamazaki, Victoria

; APPLICANT: Chen, Rui-hong

; APPLICANT: Wang, Zhiwei

; APPLICANT: Wang, Dunrui

; APPLICANT: Yang, Yonghong

; APPLICANT: Wehrman, Tom

; APPLICANT: Ghosh, Reena

; APPLICANT: Drmanac, Radoje T.

; TITLE OF INVENTION: No. 6783969el Nucleic Acids and

; FILE REFERENCE: 803

; CURRENT APPLICATION NUMBER: US/09/799,451

; CURRENT FILING DATE: 2001-03-05

; NUMBER OF SEQ ID NOS: 948

; SOFTWARE: pt_FL_genes Version 2.0

; SEQ ID NO 528

; LENGTH: 6151

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (263)..(3133)

; US-09-799-451-528

Alignment Scores:

Pred. No.: 4,85e-07 Length: 6151

Score: 169.50 Matches: 118

Percent Similarity: 36.28% Conservative: 108

Best Local Similarity: 18.94% Mismatches: 248

Query Match: 5.15% Indels: 149

DB: 4 Gaps: 22

US-10-736-868-2 (1-643) x US-09-799-451-528 (1-6151)

Qy 36 LeuValGlyLysSerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGluLeu 55

Db 632 CTGACTGGCTCATCCCATCATCACCACCACAGGTCCTCCATG---TTGAGGCGAGTA 688

Qy 56 MetGlyValGlnPheValAspAlaLeuLeuLysLysGlyGlnMetGluMetAlaLysGly 75

Db 689 AGAGACACACAATGTTAGATCTTCAGCCCGAGCTGAAAGAACTGCAGAGAGAGAAATGAC 748

Qy 76 AlaPheLysThrGlnLeuGluValLeuLysValHisProAspGlnPheAspLysTyr 95

Db 749 CTCCTCCGGAAGAGCTAGACATCAGCAGCAGCAATTTGGGATCTTCCATGAACAGATT 808

Qy 96 LysLysLeuLysValAspAspLeuAlaAlaAspAlaValMetGlnGlnAlaGluMetAla 115

Db 809 AAGACTTTCTGGAGTCTGAGCTTAAGAAGAGAGAGAGTCTTGAGGAAAGAGAGCGAGCG 868

Qy 116 LysLeuGlnProLysSerGlyAsnAlaPheLeuLeuMetLeuAsnGlyAsnGlyLeuPro 135

Db 869 CGGATGCTGCTCCTCAGGAGCAGATGAGGTTTCCCATGAAGAAATTCAGCAGCTACAG 928

Qy 136 IleGlySerSerIleArgGlyLeuGluAspAlaIleArgThrGlnArgMetGluAsn 155

Db 929 TTG-----ACATCCAGGCCCTTCAGATGAGCTGCGAACCCAGAGAGACTCAACCAC 982

Qy 156 ThrAspProSerGluGlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeu 175

Db 983 CTCCTCCAGCAAGAGAGTGGCAACCGAGGCG---GAGCACTTCACCATCAGCTGACC 1039

Qy 176 ProGlyLeuValAlaAsnMetIleAlaGlyLysAsnProPheLysMetProGlnGlnMet 195

Db 1040 -----GAGGAGAACTTTAGCGGCTCCAAGCCGAG 1069

Qy 196 ArgLysAlaGlnAlaAlaProSerSerValPheGlnGlnAlaLeuAlaGlnArgAlaMet 215

Db 1070 CATGACAGCGCAGCTTAAGGAGCTGTTCTTTTGGGAAGACATTTAGAGAAATGGAGCTG 1129

Qy 216 LeuGlyLysAsnAlaProValAlaGlyGlyArgGlyGluGlnArgMetMetMetAsn 235

Db 1130 -----AGAATTGAACGCGACAAACAAACCCCTCAAT 1159

Qy 236 ArgValaAspGlnArgMetGlnGln-----ArgGluLeuGlnGluAspGluAspAsp 253

Db 1160 GCCGAGATGAGTCAATTTAAAAAATCTTCTGAGATGTTTGAAGTAAAGGCTTGCCATCC 1219

Qy 254 AspAspLeuGluAspGluAspValProArgArgArgSerSerAspGlyGluProGln--- 272

Db 1220 AAAAGCCTGGAGGATGACATGACGCAACGCGCGGATGGCAGAGGCTGAGTCTCAGGTC 1279

Qy 273 -----SerGluAlaGluHisGlnArgAsp 281

Db 1280 AGCCACTTGGAAAGTGATTTTAGATCAGAAAGAGAAAGAAACATACATCTTTAGAGAGAA 1339

Qy 282 LeuAlaArgLeuLysSerSerProArgLeuLysGluLeuLeuGlnAsnAlaGluVal 301

Db 1340 TTGACCGAAGACCACTTCAGCG-----GAGCCAGCAAGACG 1381

Qy 302 GlnSerLeuSerTyrGlnArgMetArgAspSerProLeuSerLysArgArgProLeu 321

Db 1382 AAGGCTCTCCAGACTGTCATCGAAATGAGGACACAAAAATCGCTTCA-----TTG 1432

Qy 322 AlaMetAsnAspGluAsp---GluSerAlaPheArgAlaMetGluAlaArgAlaLysLeu 340

Db 1433 GAACGAAACATTAAGGATCTTGAGGATGAGATCCAGATGTTTAAAGCAATGTTGTGCTG 1492

Qy 341 AspGlnLysSerGlnLeuValLeuGlyLeuHisGlyPheGlyGluSerAspAspAspGlu 360

Db 1493 -----AACACTGAG 1501

Qy 361 AspGluGluAspGluAsnLeuIleAspProSerGluAsnSerPheArgArgAlaProLeu 380

Db 1502 GACCGCGAAGAGAG----- 1516

Qy 381 ArgLeuSerSerGlyPheValGluLysLeuLysSerAsnAspGluLeuLysSerAlaLeu 400

Db 1517 -----ATCAAAACAAATTCAGGTTTACAAAGTCACTCC 1549

Qy 401 AspArgIleLysTyrArgValAspValGluLysTyrLeuAlaProLysProMetGlu 420

Db 1550 AAGTTTATGAAGCAACCAAGATTGATCAGCTGAAGCAGGAACCTTTCAAGAAAGAGATCGGAA 1609

Qy 421 PheAsnProLysProGlnProGlyTyrPheAlaProArgLysIleProThrArgProArg 440

Db 1610 CTT-----CTTGCCTTACAAACAAAGCTTTGAACCTCAGCAAT 1648

Qy 441 LysMetLeuProLeuLeuIleGlySerAspProLysValGlnGluGluIleArgArgHis 460

Db 1649 CAA-----AATTCAGATTGCAAGCAACACATTTGAAGTCTCAAGAG 1690

Qy 461 ProSerThrGluTrpLysIleAlaLysGluSerArgValLeuThrAsnLeuLysAsnAsn 480

Db 1691 TCACCTTACT-----GCCAAAGAACAGAGGCTGCGCATC---CTTCAGACTGAG 1735

Qy 481 ProSerLeuAlaAlaLeuPheMetAspLys-----LeuGluAsnThrLeuLys 497

Db 1736 GTAGATCGCTGAGATTACCATCGTAGAAGAAAGAAATCTTCTTCATATAAAAAACAAA 1795

Qy 498 GlyArgGlnMetLeuThrAspGluGlnLysGlyArgThrArgValLysThrIleArgAla 517

Db 1796 CAGCTACAGGACCTCACA---GAAGAGAGGG----- 1825
Qy 518 LeuProArgLeuPheGlyAlaProThrAlaLysAlaGluMetIleAspAlaLysValPhe 537
Db 1826 -----ACACTGCCCGTGAATAATCTGTGACATCAAGATATG 1861
Qy 538 GlnAspIleGluArgProIleProLeuPhePheGluProLysGlyArgHisThr 557
Db 1862 TTAGAAGTGAAGGAAGAAATAATCAATGCTTCAGAAAAGATGTAAGAACTTGCAGAA 1921
Qy 558 ArgLeuArgTrpThrGlyAlaAsnGluLysGluIleProGlyLeuGlySerArgPheIle 577
Db 1922 CAACCTAGG-----GATAAGACACAGCAACTGACCACTGAAGACAGA----- 1966
Qy 578 LeuProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSerThr----- 594
Db 1967 GTGAAGTCTTGCAGACGGATTCAGTAATACAGATACTGCACCTGGCGACGCTAGAGAA 2026
Qy 595 -----GlnGlyArgAlaArgAspGlu 601
Db 2027 GCTCTGCAGAGAGGAGAGATAATTTAGCGCTTGAAGAACAGCAGCAAGAGATGAT 2086
Qy 602 TrpAspThrMetPheLysIleProAsnAsnTrpAsnProGlyAspGluValGlyPheLys 621
Db 2087 CGGGAAGACTAGAGAGATAGATCCTTCGAAAAGAGAACAAAGACCTGAAGAGAG 2146
Qy 622 MetAsnSer 624
Db 2147 GTCAATGCT 2155

RESULT 3

US-09-949-016-5001
; Sequence 5001, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5001
; LENGTH: 5883
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-5001

Alignment Scores:
Pred. No.: 2.54e-06 Length: 5883
Score: 162.50 Matches: 136
Percent Similarity: 35.90% Conservative: 88
Best Local Similarity: 21.79% Mismatches: 213
Query Match: 4.94% Indels: 192
DB: 4 Gaps: 25

US-10-736-868-2 (1-643) x US-09-949-016-5001 (1-5883)

Qy 12 GlyPheCysIleAlaProLeuSer-----AlaGlnSerProSerThrSerAsp----- 27
Db 3860 GGCTTCTCAGCCATCCGACAGCAAGTCCAGCAGCTCACCAGGACTTCTCCGCGTGG 3919
Qy 28 AlaProGly-----AlaLeuLeuSerSerLeuValGlyLysSerHisGlnLysLeuPro 45
Db 3920 AGTCCCAGCTGCAGGACACTCAGGAGCTG-CTGCAGGAGGAGAACCCGGCAGAGCTGAGC 3978

Qy 46 LeuAlaProSerMetGluAlaLeuGlu----- 54
Db 3979 CTGAGCAACCAAGCTCAAGCAGGTGGAGAGCAGAGAAATTCCTTCGGGGAGCAGCTGGAG 4038
Qy 55 -----LeuMetGlyValGlnPhe 60
Db 4039 GAGGAGGAGGAGCCCAAGCACCACTGGAGAACGACATGCCACCCTCCATGCCAGGTG 4098
Qy 61 ValAspAlaLeuIleLysLysGlyGlnMetGluMetAlaLysGlyAlaPheLysThrGln 80
Db 4099 GCCGAC-----ATGAAAAAAGATGAGGACAGTGTGGGTGCTCTGGAACACTGCT 4149
Qy 81 LeuGluValLeuGluLysValHisProAsp----- 90
Db 4150 GAGGAGGTGAAGAGGAGCTCCAGAGGACCTGGAGGGCTGAGCCAGCGGCACGAGGAG 4209
Qy 91 -----GlnPheAspLysTyrlLysLysLeuLys-----ValAsp 101
Db 4210 AAGGTGCCCGCTTACCACAGCTGGAGAGAACCAAGACGCGGCTGCAGCAGGAGCTGGAC 4269
Qy 102 AspLeuAlaAlaAspAlaValMetGlnGlnAlaGluMetAlaLysLeuGlnProLysSer 121
Db 4270 GACCTGCTGTGACCTGGACCAACCAAGCGCGCAGAGCGCTGCAACCTGGAGAGAAAGCAG 4329
Qy 122 GlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIleProIleGly----- 137
Db 4330 AGAAG-----TTTGACCAGCTCTGGCGGAGGAGAACCACTCTCTGCCAAGTATGCA 4383
Qy 138 -----SerSer 139
Db 4384 GAGGAGCGGACCGGGCTGAGCGGAGGCCGAGAGAACCAAGGCTCTGTCTGCTG 4443
Qy 140 IleArgGlyLeuGluAspAlaIleArgThrGlnArgAspMetGluAsnThrAspProSer 159
Db 4444 GCCCGGGCCCTGGAGGAAGCCATGTGAGCAGAGACGCGGAGCTGGAGCGGCTCAAC----- 4497
Qy 160 GluGlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeuProGlyLeuVal 179
Db 4498 -----AAGCAGTTCCGCACGGAG-----ATG 4518
Qy 180 AlaAsnMetIleAlaGlyLysAsnProPhe-----LysMetProGlnGlnMetArgLysAla 198
Db 4519 GAGGACCTTATGAGCTCCAAGGATGATGTGGGCAAGAGTGTCCACGAGCTGGAGAAAGTC- 4577
Qy 199 GlnAlaAlaProSerSerValPheGlnGlnAlaLeuAlaGlnArgAlaMetLeuGlyLys 218
Db 4578 CAAGCGGGCCCT-----AGACGACGAGGTGGAGGA 4607
Qy 219 Asn-----AlaProValAlaGlyLysArgGlyGluGluGlnArgMetMetAsn 235
Db 4608 GATGAAGAGCGCAGCTGGAGAGCTGGAGGACGAGCTGCAGGCCACCGAAGATCCCAAGCT 4667
Qy 236 -ArgValAspGlnArgMetGln-----GlnArgGluLeuGlnGlu 249
Db 4668 GCGGTTGGAGGTCAACCTGCAGGCCCATGAAGGCCAGTTCGAGCGGACCTGCAGGCGCG 4727
Qy 249 uAspGluAspAspAspLeuGluAspValProArgArgSerSerSerAspGlu 269
Db 4728 GGACGACGACGAGCGAGGAGAAAGCAGCTGGTGCACACAGCTCGCGGAGATGAGGCG 4787
Qy 269 yGluProGlnSerGluAlaGluHisGlnArgArgAspLeuAlaArgLeuLysSerSe 289
Db 4788 AGAGCTGGAGGACGAGGAGGACGCGCTCGATGGCAGTGGCGCCCGGAGAGAGCTGGA 4847
Qy 289 rProArgLeuLysGluLeu----- 295
Db 4848 GATGACGCTGAAGGACCTGGAGGCTCAGCTCGACTCGGCCCAACAGAACCGGACGCAAGC 4907
Qy 296 -----LeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrlGlnArgMetArgAs 312
Db 4908 CATCAACAGCTCGGAGAGCTGCAGGCCCAAGATGAAGAGGACTGCATCGCGAGCTGATGA 4967
Qy 312 pSerProLeuSerLysArgArgProLeuAlaMetAsnAspGluAspGluSerAlaPheAr 332


```

QY 478 LysAsnAsnProSerLeuAlaLeu-----PheMetAspAspLysLeu 492
Db 2065 AAGTTCAGTCCACCATCGCGCGCTGGAGCCCAAGATTCCACAGCTGGAGGACAGGTC 2124
QY 493 GluAsnThrLeuLysGlyArgGlnMetLeuThr----- 503
Db 2125 GAGCAGGAGCCAGAGAGAAACAGCGCAGCCACCAGTCGCTGAAGCAGAAAGACAAGAG 2184
QY 504 -----AspGlu----- 505
Db 2185 CTGAAGGAATCTTGTCTGAGTGGAGGACGAGCGCAAGATGGCCGAGCAGTACAAGGAG 2244
QY 506 -----GlnLysGlyArgThrArgValLysThrIleArgAlaLeuProArgLeuPheGly 523
Db 2245 CAGCAGAGAAGGCAATGCCAGGCTCAAGCTCAAG-----AGGCGAGCTGGAG 2295
QY 524 AlaProThrAlaLysAlaGluMetIleAspAla-----LysValPheGlnAspIle 540
Db 2296 GAGCAGAGGAGAGTCCCGAGCGCATCAACGCCAACCAGGAGAGCTGCGAGCGGAGCTG 2355
QY 541 GluGluArgProIleProProLeuPhePheGluProLysGlyArgHisThrArgLeuArg 560
Db 2356 GATGAGGCC----- 2364
QY 561 TrpThrGlyAlaGln-----LysGluIleProGlyLeuGlySerArgPheIle 577
Db 2365 ---ACGGAGAGCAACGAGGCCATGGCGCGTGAGGTGAACGCACTCAAGAGCAAGCAGA 2421
QY 578 LeuProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSerThrGlnGlyArg 597
Db 2422 GGGCCCCC-----CCACAGA-AACTTCGAGTGATGCACCGCGGAGG 2465
QY 598 AlaArgAspGluTrpAspThrMetPheLysIle 608
Db 2466 AAACGAGACCTCTTCTGCTCTCTAGAGGTC 2498

RESULT 6
US-08-533-306A-3
; Sequence 3, Application US/08533306A
; Patent No. 5837457
; GENERAL INFORMATION:
; APPLICANT: Liu, Pu
; APPLICANT: Collins, Francis S.
; APPLICANT: Siciliano, Michael J.
; APPLICANT: Claxton, David
; TITLE OF INVENTION: Markers for Detection of Chromosome 16
; TITLE OF INVENTION: Rearrangements
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: USA
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/533,306A
; FILING DATE: September 25, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, DeAnn F.
; REGISTRATION NUMBER: 36683
; REFERENCE/DOCKET NUMBER: 2115--00869C0B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 641-1600
; TELEFAX: (810) 641-0270
; INFORMATION FOR SEQ ID NO: 3:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 2887 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; INDIVIDUAL ISOLATE: Sample 1
; TISSUE TYPE: Acute myelomonocytic leukemia, M4E0
; TISSUE TYPE: subtype (inv16)
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 16(inv16)(p13q22)]
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2658
; US-08-533-306A-3

Alignment Scores:
Pred. No.: 1.95e-06 Length: 2887
Score: 159.00 Matches: 150
Percent Similarity: 38.30% Conservative: 130
Best Local Similarity: 20.52% Mismatches: 230
Query Match: 4.83% Indels: 225
DB: 2 Gaps: 35

US-10-736-868-2 (1-643) x US-08-533-306A-3 (1-2887)
QY 15 IleAlaProLeuSerAlaGlnSerProSerThrSerAspAlaProGlyAlaLeuLeuSer 34
Db 769 GTGGCGTCTCCAGTTCCTCCAGCTCCAGGACCCAGGAGTTG----- 810
QY 35 SerLeuValGlyLysSerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGlu 54
Db 811 ---CTTCAAGAGAAACCCCGCAGAGGCTCAACGTGCTCTACGAAGCTGCGCCAGCTGGAG 867
QY 54 ----- 54
Db 868 GAGGAGCGGAACAGCCTGCAAGACCAGCTGGAGGAGATGGAGCCCAAGCAGAACCTG 927
QY 55 -----LeuMetGlyValGlnPheValAspAlaLeuIleLysLysGlyGln 69
Db 928 GAGCGCCACATCTCCACTCTCAACATCCAGCTCTCCGACTCGAAG---AAGAGCTGCAG 984
QY 70 -----MetGluMetAlaLysGlyAlaPheLysThrGlnLeu 81
Db 985 GACTTTTCCAGCAGCCGCTGGAAGCTCTGGAAGAGGGAAGAGGTTCCAGAAAGGAGATC 1044
QY 82 GluValLeuGluLysValHisProAspGln-----PheAspLysTyrLysLysLeu 98
Db 1045 GAGAACCTCACCAGCAGTACGAGGAGAGGCGCGCTTATGATAAATCTGGAAGAGACC 1104
QY 99 Lys-----ValAspAspLeuAlaAlaValMetGlnGlnAla 112
Db 1105 AAGAACAGGCTTCAGCAGGAGCTGGAGACCTGGTTGTTGATTGGACCAACCGCGCAA 1164
QY 113 GluMetAlaLysLeuGlnProLysSerGlyAsnAlaPheIleAspMetLeuAsnGlyAsn 132
Db 1165 CTGTTGTCAACCTGGAAAAAGAGCAGAGGAAA-----TTTCATCATGTTGTTAGCCGAG 1218
QY 133 GlyIleProIleGlySerSer----- 139
Db 1219 GAGAAAAACATCTCTTCCAAATACCGGATGAGAGGACAGAGCTGAGGCGAAGCCAGG 1278
QY 140 -----IleArgGlyLeuGluAspAlaIleArgThrGln 150
Db 1279 GAGAAAGAAACCAAGGCCCTGTCTCCCTGGCTCGGGCCCTTGAAGAGGCTTGGAGGCCAA 1338
QY 151 ArgAspMetGluAsnThrAspProSerGluGlnIleAlaLysAlaValMetAspLysPhe 170
Db 1339 GAGGAACTCGAGCGGAGCCAAAC-----AAAATGCTCAAGCCGAAATGGAA----- 1383

```



```

Db      2392 CTGAAGGAATCTTGTGCAGGTGAGGACGAGCGCAAGATGCCGAGCAGTACAAAGGAG 2451
QY      |||||
506  -----GlnLysGlyArgThrArgValLysThrIleArgAlaLeuProArgLeuPheGly 523
Db      :|||:
2452 CAGGCAGAGAAAGCAATCCAGGTCAAGCAGCTCAAG-----AGGCAGCTGGAG 2502
QY      |||||
524 AlaProThrAlaLysAlaGluMetIleAspAla-----LysValPheGlnAspIle 540
Db      :|||:
2503 GAGGCAGAGGAGGAGTCCAGCGCATCAACGCCCAACCGCAGGAAGCTGCAGCGGAGCTG 2562
QY      |||||
541 GluGluArgProIleProLeuPheGluProLysGlyArgHisThrArgLeuArg 560
Db      :|||:
2563 GATGAGGCC----- 2571
QY      |||||
561 TrpThrGlyAlaAsnGlu-----LysGluIleProGlyLeuGlySerArgPheIle 577
Db      :|||:
2572 ---ACGGAGAGCAACGAGCCATGGCGGTGAGGTGAACGCACTCAAGAGCAAGCTCAGA 2628
QY      |||||
578 LeuProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSerThrGlnGlyArg 597
Db      :|||:
2629 GGGCCCCC-----CCACAGGA-AACTTCGCAGTGATGCACCAGCGCAGG 2672
QY      |||||
598 AlaArgAspGluTrpAspThrMetPheLysIle 608
Db      :|||:
2673 AAACGAGACCTCTTTCGTCTCTTAGAAGGTC 2705

```

RESULT 8

```

US-09-949-016-1019
; Sequence 1019, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1019
; LENGTH: 4935
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-1019

```

```

Alignment Scores:
Pred. No.: 4,56e-06 Length: 4935
Score: 159.00 Matches: 159
Percent Similarity: 36.06% Conservative: 97
Best Local Similarity: 22.39% Mismatches: 260
Query Match: 4.83% Indels: 196
DB: 4 Gaps: 35

```

US-10-736-868-2 (1-643) x US-09-949-016-1019 (1-4935)

```

QY      23 ProSerThrSerAspAlaProGlyAlaLeuLeuSer---SerLeuValGlyLysSerHis 41
Db      |||||
484 CCGAGTACCTACGCAAGATGGCGAGCTGGAGGAGGTGACTCTGGACGGGAAGCCTCTT 543
QY      |||||
42 GlnLysLeuProLeuAlaProSerMetGluAlaLeuGluLeuMetGlyValGlnPheVal 61
Db      :|||:
544 CAGGCGCTGCGGTGACCGACCTGAAGCCGCACTGGAGCAGCGAGGC----- 591
QY      |||||
62 AspAlaLeuIleLysLysGlyGln-----MetGluMetAlaLysGly----- 75
Db      |||||
592 -----CTAGCCAAAGAGCGGCGCAAGAGAGTGCCTCTGGTCAAGCGGCTCAAAAGGGGCTCTA 645

```

```

QY      76 -----AlaPheLysThrGlnLeuGlu 82
Db      |||||
646 ATGTAGAAAATTTACAGAAACACTCAACACCCCATGTGCTATCCAGCCAAATTTCCAG 705
QY      |||||
83 ValLeuGluLysValHisProAspGlnPheAspLysThrLysLysLeuLysValAspAsp 102
Db      :|||:
706 ATTGTGAGGAAATGAGCCAGAACAGATTTCATAAAACAGTATCTCGAAAAGCAGCAGGAG 765
QY      |||||
103 LeuAlaAlaAspAlaValMetGlnGlnAla---GluMetAlaLysLeuGlnProLysSer 121
Db      :|||:
766 CTACTTAGCGAGCGTCTGGAACGTGAAGAGCTCGAAGAGCTGCAGAACTTTGAAGAAGCTTCA 825
QY      |||||
122 GlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIle-----Pro 135
Db      :|||:
826 GCTGAGTCGGAGCAGCAGATGATCCATCTCGAGGGAGTGGCTTCCTGCTGCTCCTGAC 885
QY      |||||
136 IleGlySerSerIleArgGlyLeuGluAspAlaIleArgThrGlnArgAspMetGluAsn 155
Db      :|||:
886 TTTGAGAGCAGCCTGGAGAGACCAAGCTGGAGCTCAGCAGACATTCGCCCAAAAAGC 945
QY      |||||
156 ThrAspProSerGluGlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeu 175
Db      :|||:
946 TCCTCAATTTCTGAAGAG----- 963
QY      |||||
176 ProGlyLeuValAlaAsnMetIleAlaGlyLysAsnProPheLysMetProGlnGlnMet 195
Db      :|||:
964 -----AAGTGCATCTGATGATGAGAAACCA----- 990
QY      |||||
196 ArgLysAlaGlnAlaAlaProSerSerValPheGlnGlnAlaLeuAlaGlnArgAlaMet 215
Db      :|||:
991 AGGAAAGGAGAAAGACGATCATCTAGGTC---AGACAGGCAAGAGCAGCTAAAGTGCT 1047
QY      |||||
216 LeuGlyLysAsnAlaProValAlaGlyGlyArgGly-GluGlu-----GlnArgMetMe 233
Db      :|||:
1048 GAGGTC---AGCCAACTCTGTGA---GGAGGAAGAGATCAAGAAACACCTTCCAGAAACCT 1103
QY      |||||
233 tMetAsnArgValAspGlnArgMet-----GlnGlnArgGluLeuGlnGluAspGln 251
Db      :|||:
1104 AAGGTCAGAGCAGATCGAAATTTGAAACAGAGGAGGAAGAGAGGAGGAGGAGAGA 1163
QY      |||||
251 uAspAspAspAspLeuGluAspGluAspValProArgArgArgSerSerAspGlyGluPr 271
Db      :|||:
1164 GGAAGAGATGATGAAGAAGAGGAA-----GGTGATGATGAGGG 1202
QY      |||||
271 oGlnSerGluAlaGluHisGlnArgArgAspLeuAlaArgArgLeuLysSerSerProAr 291
Db      :|||:
1203 ACAAAATCTAGGAG-----GCACCAAT 1226
QY      |||||
291 gLeuLysGluLeuLeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrGlnArgMetAr 311
Db      :|||:
1227 CCTGAAGAGTTTAAGGAGAGAGGGGAA-----CAGATACCTAGAGTAAA 1271
QY      |||||
311 gAspSerProLeuSerLysArgArgProLeuAlaMetAsnAspGluAspGlu----- 328
Db      :|||:
1272 ACCAGAGGATGATGATGATGAGACCCCAAAACAGATCCCAAGAGGAGGTGTTAGA 1331
QY      |||||
329 -----SerAlaPheArgAlaMetGluAlaArgAlaLysLeuAspGlnLysSerGlnLe 346
Db      :|||:
1332 GAGAGGAGGAGATTTCAGAGATCCAGGAGAGGCTAGA-----AAAAGTCATCT 1382
QY      |||||
346 uValLeuGlyLeuHisGlyPheGlyGluSerAspAspGluAspGluGluAspGluAs 366
Db      :|||:
1383 GGCC-----AGACAGCAGCAGGAGAGAGAAATGAA 1412
QY      |||||
366 nLeuIleAspProSerGluAsnSerPheArgAlaProLeuArgLeuSerSerGlyPh 386
Db      :|||:
1413 AACAACTCTCCCTTGAGGAGGAGAAAGAGAGAA-----ATAAAATCTTCAAGGCTT 1466
QY      |||||
386 eValGluLysLysSer-----AsnAspGluLeuLysSerAl 399
Db      :|||:
1467 AAAGGAAATAATCGAAGTCTCTTCCCTCTCGACTGAGTGAAGATCAAGAGAGGCTC 1526

```

```
QY 399 aLeuAspArgIleIysTyrArgValAspAspValGluIysTyrLeuAlaProLysProme 419
Db 1527 ACTTGTCGCTGCCAGCAAACTGCCAGCGAGGAG------ACTCCTCCACCTTT 1580
QY 419 t-----GluPheAsnProLysProGlnProGlyTyrPheAlaProArgLysil 435
Db 1581 ACTAACAAAGGAAGCATCTTCTCCACCACTCTCCACAGCTCCACAGCTCCATAGCAAGAAAT 1640
QY 435 e-----ProThrArgPro----- 439
Db 1641 AGAGCCCATGGAAGGCCAGCCGCCCTGCTCTCACTTCAGTTATCTCTCCTAATACAGA 1700
QY 440 -----ArgLysMetLeu----- 443
Db 1701 TGTGACACCGAGGAGTATTAGTATCTCAGCATCTGTCAGTGTGGTAGAGGCTGTCT 1760
QY 444 -ProLeuLeuIleGlySerAsp-ProLys-----ValGlnGluGluIleArgArgH 460
Db 1761 TCCTTTGTCAAGTCTTCAGACCAACAAAGCAGAAATCTCCAGCAGAGAAAGTGCAGAGGA 1820
QY 460 isProSerThrGluTrp-----LysIleAlaLysGluSerArgValLeuT 475
Db 1821 GAGTGTCTGCTGCTGTTTTCAGAAAGCACACTGGCTGACTACTCAGCCCAAGAGGATCT 1880
QY 475 hrAsnLeuLysAsn-----AsnProSerLeuAlaAlaLeu-----P 487
Db 1881 TGAACCTGATCGACAGATCTGCTCAGCCCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1940
QY 487 heMetAspAspLysLeuGluAsnThrLeuLysGlyArgGlnMetLeuThrAspGluGlnL 507
Db 1941 GGCCAAAGGATCACTGAAGATGT-CTGAAA-----CAGCCATCTTTGGACAGAGG 1993
QY 507 ysGlyArgThrArgValIleThrIleArgAlaLeuProArgLeuPheGlyAlaProThrA 527
Db 1994 AAGCAGAAAGAGTCTCTATACCTCTCTCCCAAGCCACAGATTGAAAACAGTCAGCTGATT 2053
QY 527 lalysAlaGluMetIleAspAlaLysValPheGlnAspIleGluIleArgProIleProp 547
Db 2054 CATCTCTAGCCGGTCTCTCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2113
QY 547 roLeuPhePheGluProLysGly-----ArgHisThrArgLeuArg----- 560
Db 2114 CT-----GACAGTTCAGTTCCTCGGTCTCTATTCACCGCTCAGATCCAGCAGAGAG 2164
QY 561 -----TrpThrGlyAlaAsnGluLysGluIleProGlyLeuGlySerArgp 576
Db 2165 ATGTAGCCAGGACGCTACTCATGCCAACCTCGTGTAGACCAAGATGGGCTCCAGAT 2224
QY 576 heIleLeuProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSer----- 593
Db 2225 CAACATCAGATCCAGATCAAGGTCAGTTTCAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 2284
QY 594 -----ThrGlnGlyArgAlaArgAspGluTrpAspThrMetPheLysIleProA 610
Db 2285 GAAAATCTCTGAGCCCTGGAGTCTCCAGGGACAGC---AGCACCGATCTACTGAAACCA 2341
QY 610 snAsnTrpAsnProGlyAspGluVal 618
Db 2342 AAGATCCCTCTCTGCTGAGGAGTT 2367

RESULT 9
US-09-949-016-5054
; Sequence 5054, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CU001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
```

```
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5054
; LENGTH: 4935
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-5054

Alignment Scores:
Pred. No.: 4,56e-06 Length: 4935
Score: 159.00 Matches: 159
Percent Similarity: 36.06% Conservative: 97
Best Local Similarity: 22.39% Mismatches: 260
Query Match: 4.83% Indels: 196
Gaps: 35
DB:

US-10-736-868-2 (1-643) x US-09-949-016-5054 (1-4935)
QY 23 ProSerThrSerAspAlaProGlyAlaLeuLeuSer---SerLeuValGlyLysSerHis 41
Db 485 CCGAGTACCTCAGCAGATGCGGAGGTGAGGAGTGAAGTCTCTGACGGGAGCCCTCTT 544
QY 42 GlnLysLeuProLeuAlaProSerMetGluAlaLeuGluLeuMetGlyValGlnPheVal 61
Db 545 CAGGCGCTGCGGTGACCGACCTGAAAGCGCCACTGAGCGCAGCGAGGC----- 592
QY 62 AspAlaLeuIleLysLysGlyGln-----MetGluMetAlaLysGly----- 75
Db 593 -----CTAGCCAAAGCGCGGAGAGAGTGCCCTGTCAGAGCGGTCAAGGGGTCTTA 646
QY 76 -----AlaPheLysThrGlnLeuGlu 82
Db 647 ATCTAGAAAAATTACAGAAACACTCAACACCCCATGCTGCATTCACGCCAATTCCTCAG 706
QY 83 ValLeuGluLysValHisProAspGlnPheAspLysTyrLysLysLysLysValAspAsp 102
Db 707 ATTGTGTGAGGAATGAGCCAGAACAGTCTTCAATAAAACAGATATCTGAAAAGCAGCAGGAG 766
QY 103 LeuAlaAlaAspAlaValMetGlnGlnAla---GluMetAlaLysLeuGlnProLysSer 121
Db 767 CTATTAGCAGCGCTCTGGAACCTGGAAGCTCGAAGAGCTGCAGACTTTGAGAAGCTTCA 826
QY 122 GlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIle-----Pro 135
Db 827 GCTGAGTCGAGGAGCAGATGATCCATCTCCTGAGGAGTGGCTTCCCTGCTGCTCTGAC 886
QY 136 IleGlySerSerIleArgGlyLeuGluAspAlaIleArgThrGlnArgAspMetGluAsn 155
Db 887 TTTTCAGAGCAGCTGAGAGACCCAGAGCTGGAGCTCAGCAGACATTCGCCCAAGAAAGC 946
QY 156 ThrAspProSerGluGlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeu 175
Db 947 TCCTCAATTTCTGAAGAG----- 964
QY 176 ProGlyLeuValAlaAsnMetIleAlaGlyLysAsnProPheLysMetProGlnGlnMet 195
Db 965 -----AAAGGTGACTCTGATGATGAGAAACCA----- 991
QY 196 ArgLysAlaGlnAlaAlaProSerSerValPheGlnGlnAlaLeuAlaGlnArgAlaMet 215
Db 992 AGGAAAGGAGAAAGACGATCATCTAGGGTC---AGACAGCGCAAGACGACGCTAAACTGTCT 1048
QY 216 LeuGlyLysAsnAlaProValAlaGlyGlyArgGly-GluGlu-----GlnArgMetMe 233
Db 1049 GAGGGC---AGCCAACTGCTGA-GGAGGAAGAGGATCAAGAAACACCTTCCAGAAACCT 1104
QY 233 tMetAsnArgValAspGlnArgMet-----GlnGlnArgGluGluGlnGluAspG1 251
Db 1105 AAGGGTCAGCAGCATCGAAATTTGAAAACAGAGGAGGAAGAGAGGAGGAGGAGGAGGA 1164
```

Qy	251	uasPaspAspAspLeuGluAspGluAspValProArgArgSerSerAspGlyGluPr	271
Db	1165	GAAGAAGATGATGAAGAAGGAA-----GGTGAATGATGAGGG	1203
Qy	271	oGlnSerGluAlaGluHisGlnArgArgAspLeuAlaArgLeuLysSerSerProAr	291
Db	1204	ACAAAATCTAGGAG-----GCACCAAT	1227
Qy	291	gLeuLysGluLeuLeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrGlnArgMetAr	311
Db	1228	CCTGAAGAGTTTAAGGAAGAAGGGAA-----GAGATACCTAGAGTAAA	1272
Qy	311	gAspSerProLeuSerLysArgArgProLeuAlaMetAsnAspGluAspGlu-----	328
Db	1273	ACCAGAGGAGATGATGGATGAGAGACCCAAACAAAGATCCCAGGACAGGAGGTGTAGA	1332
Qy	329	-----SerAlaPheArgAlaMetGluAlaArgAlaLysLeuAspGlnLysSerGlnLe	346
Db	1333	GAGAGGAGGAGATTACAAGATCCAGGAAGAGCTAGA-----AAAAGTCATCT	1383
Qy	346	uValLeuGlyLeuHisGlyPheGlyGluSerAspAspGluAspGluAluAspGluAs	366
Db	1384	GGCC-----AGACAGCAGCAGGAGGAAGAAATCAA	1413
Qy	366	nLeuLeaAspProSerGluAsnSerPheArgArgAlaProLeuArgLeuSerSerGlyPh	386
Db	1414	AACAACATCTCCCTTGAGGAGGAAGAAAGAA-----ATAAAATCTTCACAAAGCCTT	1467
Qy	386	eValGluLysLeuLysSer-----AsnAspGluLeuLysSerSerAl	399
Db	1468	AAAGGAAAAATCGAAGTCTCCTCCCTCGACTGACTGAAGATCGAAGAAGGCCTC	1527
Qy	399	aLeuAspArgIleLysTyrArgValAspAspValGluLysTyrLeuAlaProLysProMe	419
Db	1528	ACTTGTAGCGCTGCAGACGAAACTGCCGCGAGGAG-----ACTCCTCCACCTTT	1581
Qy	419	t-----GluPheAsnProLysProGlnProGlyTyrPheAlaProArgLysIl	435
Db	1582	ACTAACAAAGGAAGCATCTCTCCACACCTCATCCAGTCCATAGCGAAGAGAAT	1641
Qy	435	e-----ProThrArgPro-----	439
Db	1642	AGAGCCCATGGAAGGCCAGCCCCCTGTCTCTCATTCAGTTATCTCCTCCTAATACAGA	1701
Qy	440	-----ArgLysMetLeu-----	443
Db	1702	TGCTGACACAGGAGGAGTATTAGTATCTCAGCATPACTGTCCAGTTGTAGGAGGCCTGTC	1761
Qy	444	-ProLeuLeuIleGlySerAsp-ProLys-----ValGlnGluGluIleArgArgH	460
Db	1762	TCCTTTGTCAAGTCTTCAGACACCAAGCGAATCTCCAGCAGAGAAGATGCCAGGGA	1821
Qy	460	isProSerThrGluTrp-----LysIleAlaLysGluSerArgValLeuT	475
Db	1822	GAGTGTCTCGCTCTGTGTCAGAAAAACACACTGGCTGACTACTCAGCCCCAGAGGATCT	1881
Qy	475	hrAsnLeuLysAsn-----AsnProSerLeuAlaAlaLeu-----P	487
Db	1882	TGAACCTTGAGTCACACAGATCTGCTCAGCCCTCCCTCTAAAAATTGAGGAATTAGCCACT	1941
Qy	487	heMetAspAspLysLeuGluAsnThrLeuLysGlyArgGlnMetLeuThrAspGluGlnL	507
Db	1942	GGCCAAAGGAATCACTGAGAGATGT-CTGAAA-----CAGCCATCTTCGACAGAGG	1994
Qy	507	ysGlyArgThrArgValLysThrIleArgAlaLeuProArgLeuPheGlyAlaProThrA	527
Db	1995	AAGGCAGACAGACTTCTCATACCTTCTCCAGGCCACACAGATTGAAACAGTCAAGT	2054
Qy	527	lalsAlaGluMetIleAspAlaLysValPheGlnAspIleGluGluArgProIleProp	547
Db	2055	CATCTCTACCGGCTCCTCACTATCTTCCTCCCTAGATCAAGATCTCGTCTC	2114

```

Qy 547 r0LeuPhePheGluProLysGly-----ArgHisThrArgLeuArg----- 560
Db 2115 CT-----GACAGTTCAGGTCTTCGGTCTCATTCACCGCTCAGATCAACGACGAG 2165
Qy 561 -----TriThrGlyAlaAsnGluLysGluIleProGlyLeuGlySerArgP 576
Db 2166 ATGTAGCCAGGCACGTACTCATGCCAACCTCGTGGTAGACCCACAGATGGGCTCCAGAT 2225
Qy 576 heileLeuProSerLeuAspProThrMetProAlaLeuAenThrAlaPheSer----- 593
Db 2226 CAACATCAGAGTCCAGATCAAGGTCACGTTTCAGTTCCTCGTTTCAGCATCAAGCAACAGCA 2285
Qy 594 -----ThrGlnGlyArgAlaArgAspGluTrpAspThrMetPheLysIleProA 610
Db 2286 GAAATCTCTGAGCCCTCGAGTCTCCAGGACAGC---AGCACCGCTATATCTGAAACCA 2342
Qy 610 snAenTrpAenProGlyAspGluVal 618
Db 2343 AAGATCCCTCTTCGTTCAGAGGTT 2368

RESULT 10
US-09-917-254-40
; Sequence 40, Application US/09917254
; Patent No. 6703204
; GENERAL INFORMATION:
; APPLICANT: Mutter, George
; APPLICANT: Baak, Jan
; TITLE OF INVENTION: Prognostic Classification of Breast Cancer
; FILE REFERENCE: B0801/7224(JRV)
; CURRENT APPLICATION NUMBER: US/09/917,254
; CURRENT FILING DATE: 2001-07-22
; PRIOR APPLICATION NUMBER: US 60/222,093
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 40
; LENGTH: 5574
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-09-917-254-40

Alignment Scores:
Pred. No.: 5.53e-06 Length: 5574
Score: 159.00 Matches: 137
Percent Similarity: 39.15% Conservative: 121
Best Local Similarity: 20.79% Mismatches: 212
Query Match: 4.83% Indels: 192
DB: Gaps: 32

US-10-736-868-2 (1-643) x US-09-917-254-40 (1-5574)

Qy 15 IleAlaProLeuSerAlaGlnSerProSerThrSerAspAlaProGlyAlaLeuLeuSer 34
Db 3583 GTGGGTCCCTCAGTTCTCCAGCTCCAGGACCCAGGAGCTG----- 3624
Qy 35 SerLeuValGlyLysSerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGlu 54
Db 3625 ---CTTCAAGAAAGAACCCGGCAGAGACTCAACGTGTCTACGAAGCTGCGCCAGCTGGAG 3681
Qy 54 ----- 54
Db 3682 GAGGAGCGGAACAGCCTGCAAGACCAGCTGGACGAGAGATGGAGGCCAAGCAGAACCTG 3741
Qy 55 -----LeuMetGlyValGlnPheValAspAlaLeuIleLysGlyGln 69
Db 3742 GAGGCGCACATCTCCACTCTCAACTCCAGCTCTCCGACTCCGAAG---AAGAGCTGCAG 3798
Qy 70 -----MetGluMetAlaLysGlyAlaPheLysThrGlnLeu 81
Db 3799 GACTTTGCCAGCACCGTGGAGTCTCTGGAAAGAGGGGAAGAGAGGTTCAGAAAGGAGATC 3858
Qy 82 GluValLeuGluLysValHisProAspGln-----PheAspLysTyrLysLysLeu 98

```

3859	Db	GAGAACCTCACCCAGCAGTACGAGGAGAAAGCGCGCGCTTTATGATAACTGGAAAAAGACC	3918	
	Qy	Lys	-----ValAspAspLeuAlaAlaAspAlaValMetGlnGlnAla	112
	Db			
3919	Db	AAGAACAGGCTTCAGCAGGAGCTGGACGACCTGGTGTGATTTGGACACACCGCGCAA	3978	
	Qy	113	GlutMetAlaLysLeuGlnProLysSerGlyAsnAlaPheIleAspMetLeuAsnGlyAsn	132
	Db			
3979	Db	CTCGTGTCAACCTGGAAAAAGACAGAGAAA-----TTTGATCAGTTGTTAGCCGAG	4032	
	Qy	133	GlyIleProIleGlySerSer-----	139
	Db			
4033	Db	GAGAAAAACATCTCTTCCAAATACGCGGATCGAGGGACAGAGCTGAGGCAGAGCCAGG	4092	
	Qy	140	-----IleArgGlyLeuGluAspAlaIleArgThrGln	150
	Db			
4093	Db	GAGAAGGAACCAAGGCGCTGTCTCGCTCGGGCCCTTGAAGAGGCTTTGGAAGCCAAA	4152	
	Qy	151	ArgAspMetGluAsnThrAspProSerGluGlnIleAlaLysAlaValMetAspLysPhe	170
	Db			
4153	Db	GAGGAACCTCAGCGGACCAAC-----AAATGCTCAAGCCGAAATGGAA-----	4197	
	Qy	171	GlnThrGlnIleLeuProGlyLeuValAlaAsnMetIleAlaGlyLysAsnProPhe---	189
	Db			
4198	Db	-----GACCTGTGCTAGCTCCAGGATGACGTGGGC	4227	
	Qy	190	LysMetProGlnMetArgLysAlaGlnAlaAlaProSerSerValPheGlnGlnAla	209
	Db			
4228	Db	AAGAACCTCATGAGCTGGAGAGTCCAAAGCGGCCCTTGAGAGCCCNAGATGGAGGAGATG	4287	
	Qy	210	LeuAlaGlnArgAlaMetLeuGlyLysAsnAlaProValAlaGlyArgGlyGluGlu	229
	Db			
4288	Db	AAGACGCAGCTGGAAGACTGGAGGACGAGCTCAAGCCACG-----GAGGAC	4335	
	Qy	230	GlnArgMetMetMetAsnArgValAspGlnArgMetGln-----Gln	243
	Db			
4336	Db	GCCAAACTG-----CGCTGGAGTCAACATGCAGCGCTCAAGGCGCCAGTTTCGA	4386	
	Qy	244	ArgGluLeuGlnGluAspGluAspAspAspLeuGluAspGluAspValProArg	263
	Db			
4387	Db	AGGGATCTCAAGCCCGGACGACGAAATGAG-----GAGAAGAGGAGGCAACTGCAG	4440	
	Qy	264	ArgArgSerSerAspGlyGluProGlnInsSerGluAlaGluHisGlnArgArgAspLeuAla	283
	Db			
4441	Db	AGACAGTTCACGATGATGACCGAAGTGGAGACGACGCGAAAGCAACGTGCCCTGGCA	4500	
	Qy	284	ArgArgLeuLysSerSerProArgLeuLysGluLeuLeuGlnAsnAlaGluValGlnSer	303
	Db			
4501	Db	GCTGCAGCAAGAAG-----AAGCTGGAAGGGGACCTGAAAGACCTGGAGCTTCAGGCC	4554	
	Qy	304	LeuLeuSerTyrGlnArgMetArgAspSerProLeuSerLysArgArgProLeu-----	321
	Db			
4555	Db	---GACTCTCCCATCAAGGGGAGGGAGGAAGCCATCAAGACGCTACCAACACTGCAGGCT	4611	
	Qy	322	AlaMetAsnAsp-----GluAspGluSerAlaPheArg---AlaMetGlu	335
	Db			
4612	Db	CAGATGAAGACTTTCAAAGAGAGCTCGAAGATGCCCGCTGCCCGACAGATGAGATCTTT	4671	
	Qy	336	AlaArgAlaLysLeuAspGlnLysSerGlnLeuValLeuGlyLeuHisGlyPheGlyGlu	355
	Db			
4672	Db	GCCACAGCCAAAGAGAAATGAGAAAGAACCAAGAGCTTG-----GAA	4713	
	Qy	356	SerAspAspAspGluAspGluGluAspGluAsnLeuIleAspProSerGluAsnSerPhe	375
	Db			
4714	Db	GCAGACCTCATGCAGCTACAAGAGGAC-----CTCCCGCGCGCTGAGAGGGCTTCGC	4764	
	Qy	376	ArgArgAlaProLeu-----ArgLeuSerSerGlyPheValGluLysLeuLysSer	392
	Db			
4765	Db	AAACAACGGACCTTCGAGAAGGAGGAAGTGGCAGAGGAGCTGGCCAGTAGACCTGTTCGGGA	4824	
	Qy	393	AsnAspGluLeuLysSerAlaLeuAspArgIleLysTyrArgValAspAspValGluLys	412
	Db			
4825	Db	AGGAACCACTCCAGGACGAGAAAGCGCGCTGGAGCGCGCGATTCGCCCTGAGAGGAG	4884	
	Qy	4825	AGGAACCACTCCAGGACGAGAAAGCGCGCTGGAGCGCGCGATTCGCCCTGAGAGGAG	4884

Qy 413 TyrIeuAlaProLysProMetGluPheAsnProLysProGlnProGlyTyrPheAlaPro 432
 Db 4885 GAGCTGAGGAGGAGGAGGG-CAACATGGAGGCCATGAGCGACCGGGTCC-GCAAGAGCCA 4942
 Qy 433 ArgIysIleProThr-----ArgProAspGlyLys 441
 Db 4943 CACAGCAGGCGGAGCAGCTCAGCAACGAGCTGGCCACAGAGCCGACGCGCCCAAGAAGA 5002
 Qy 442 Met---LeuProLeuLeuIleGlySerAspProLysValGlnGluIleArgArgHis 460
 Db 5003 ATGAGAGTGCC-----GGCAGCAGCTCGAGCGCGCAGAACAGAGGACTCCGGAGCA 5053
 Qy 461 ProSerThrGluThrIpsIleAlaLysGluSerArgValLeuThrAsnLeuLysAsnAsn 480
 Db 5054 AGCTCCACGAGATGGAGGGG-GCGCTCAAGTCC-----AAGTTCAAG 5094
 Qy 481 ProSerLeuAlaLeu-----PheMetAspAspLysLeuGluAsnThr 495
 Db 5095 TCCACCATCGGGCGCTGGAGGCCAAGATTGCACAGCTGGAGGAGCAGGTCTGAGCAGGAG 5154
 Qy 496 LeuLysGlyArgGlnMetLeuThr-----ArgGlu-----Gln 506
 Db 5155 GCCAGAGAGAAACAGCGCGCCACCAAGTCGCTGAAGCAGAAACAGCAAGAGCTGAAGGA 5214
 Qy 504 -----ArgGlu----- 504
 Db 5215 ATCTTGCTGCAGGTGGAGCAGCAGCGCAAGATGGCCGAGCAGTACAGGAGCAGGCGAGAG 5274
 Qy 507 LysGlyArgThrArgValLysThrIleArgAlaLeuProArgLeuPheGlyAlaProThr 526
 Db 5275 AAGGCAATGCCAGGTTCAGCAGCTCAAG-----AGGCAGCTGGAGGAGGCGAGAG 5325
 Qy 527 AlaLysAlaGluMetIleAspAla-----LysValPheGlnAspIleGluGlu 542
 Db 5326 GAGGAGTCCCGGCATCAACGCGCAACCGCAGGAAGTGCAGCGGGAGCTGGATGAG 5382

RESULT 11
 US-09-949-016-1240
 ; Sequence 1240, Application US/09949016
 ; Patent No. 6812339
 ; GENERAL INFORMATION:
 ; APPLICANT: VENTER, J. Craig et al.
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 ; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THERE
 ; FILE REFERENCE: CLO01307
 ; CURRENT APPLICATION NUMBER: US/09/949,016
 ; CURRENT FILING DATE: 2000-04-14
 ; PRIOR APPLICATION NUMBER: 60/241,755
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/237,768
 ; PRIOR FILING DATE: 2000-10-03
 ; PRIOR APPLICATION NUMBER: 60/231,498
 ; PRIOR FILING DATE: 2000-09-08
 ; NUMBER OF SEQ ID NOS: 207012
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1240
 ; LENGTH: 8661
 ; TYPE: DNA
 ; ORGANISM: Human
 ; US-09-949-016-1240

RESULT 11

```

RES001.11
US-09-949-016-1240
; Sequence 1240, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1240
; LENGTH: 6861
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-1240

```

Alignment Scores:	
Score, No.:	7.68e-06
Pred.	159.00
Percent Similarity:	39.11%
Best Local Similarity:	20.7%
Query Match:	4.83%
DB:	4
Length:	6861
Matches:	137
Conservative:	121
Mismatches:	212
Indels:	192
Gaps:	32

US-10-736-868-2 (1-643) x US-09-949-016-1240 (1-6861)

Qv 15 IleAlaProLeuSerAlaGlnSerProSerThrSerAspAlaProGlyAlaLeuLeuSer 34


```
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1241
; LENGTH: 6861
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-1241

Alignment Scores:
Pred. No.:      7.68e-06       Length:      6861
Score:         159.00        Matches:     137
Percent Similarity: 39.15%    Conservative: 121
Best Local Similarity: 20.79% Mismatches:   212
Query Match:      4.83%      Indels:      192
DB:              Gaps:      32

US-10-736-868-2 (1-643) x US-09-949-016-1241 (1-6861)
QY      15 ILEAlaProLeuSerAlaGlnSerProSerThrSerAspAlaProGlyAlaLeuLeuSer 34
      ::::| | | | | | | | | | | | | | | | : : : 
Db      4016 GTGGCGTCCCTCAGTTCCAGCTCCAGGCACCCAGGAGCTG----- 4057

QY      35 SerLeuValGlyLysSerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGlu 54
      |||| : : : : : | | | | | | | | | | : : : 
Db      4058 ---CTTCAGAAGAAGAAACCGGCAGAGCTCAACGTGTCTACGAAGCTGCCAGCTGGAG 4114

QY      54 ----- 54

Db      4115 GAGGAGCGGAACAGCTTGCAAGACCAGCTCGACGAGGAGATGGAGGCCAAGACAACCTG 4174

QY      55 -----LeuMetGlyValGlnPheValAspAlaLeuIleLysLysGlyGln 69
      :: : : : : | | | | | | | | | | : : : 
Db      4175 GAGGCCACATCTCCACTCTCAACATCCAGCTCTCCGACTCGAAG---AGAAGCTGCAG 4231

QY      70 -----MetGluMetAlalysGlyAlaPheLysThrGlnLeu 81
      |||| : : : : | | | | | | | | | | : : : 
Db      4232 GACTTTGCCAGCACCGTGAAGAGCTCTGGAAGGGGAAGAAGAGGTTCAGAAAGGAGATC 4291

QY      82 GluValLeuGluLysValHisProAspGln-----PheAspLysTyrlLysLysLeu 98
      |||| : : : : : | | | | | | | | | | : : : 
Db      4292 GAGAACCTCACCCAGCATACGAGGAAAGCGCGCGCTTATGATAAACCTGSAANAAGACC 4351

QY      99 Lys-----ValAspSpleuAlaAlaAspAlaValMetGlnGlnAla 112
      |||| : : : : | | | | | | | | | | : : : 
Db      4352 AAGAACAGGCTTCAGCAGGAGCTGGACACACTGTTGTTGATTGGACAACCCAGCGCAA 4411

QY      113 GluMetAlalysLeuGlnProLysSerGlyAsnAlaPheIleAspMetLeuAsnGlyAsn 132
      :::::: : : : : | | | | | | | | | | : : : 
Db      4412 CTCGTGTCCAACCTTGAAAAAGACGAGGAAA-----TTTGATCAGTTGTTAGCCGAG 4465

QY      133 GlyIleProIleGlySerSer----- 139

Db      4466 GAGAAAAACATCTCTTCCAATAACGCGGATGAGAGGCACAGACTGAGCGAAGCCOAGG 4525

QY      140 -----IleArgGlyLeuGluAspAlaIleArgThrGln 150
      :::::: : : : : | | | | | | | | | | : : : 
Db      4526 GAGNAGAAACCAAGGCCCTGTCCCTGGCTCGGGCCCTTGAGAGGCCCTTGGAGGCCAAA 4585

QY      151 ArgAspMetGluAsnThrAspProSerGluGlnIleAlalysAlaValMetAspLysPhe 170
      :::::: : : : : | | | | | | | | | | : : : 
Db      4586 GAGGAACCTCGAGCGGACCAAC-----AAAATGCTCAAGCGCAAAATGGAA---- 4630

QY      171 GlnThrGlnIleLeuProGlyLeuValAlaAsnMethIleAlaglyLysAsnProPhe--- 189
      :::::: : : : : | | | | | | | | | | : : : 
Db      4631 -----GACCTGCTCAGCTCCAAGGATGAGCTGGGC 4660

QY      190 LysMetProGlnGlnMetArgLyseAlaGlnAlaProSerSerValPheGlnGlnAla 209
      :::::: : : : : | | | | | | | | | | : : : 
Db      4661 AAGAACTCATGAGCTGGAGAGCTTCACAGCGGGCCCTGGAGACCCCAGATGGAGGAGATG 4720

QY      210 LeuAlaGlnArgAlaMetLeuGlyLysAsnAlaProValAlaGlyGlyArGglyGluGlu 229
```



```
Db 5708 AAAGCAATGCCAGGTCAAGCAGCTCAAG-----AGCAGCTGGAGGAGCAGAG 5758
Qy 527 AlaLysAlaGluMetIleAspAla-----LysValPheGlnAspIleGluGlu 542
Db 5759 GAGGAGTCCAGCGCATCAACCCCAACCCAGCAGGAGCTGCACCGGGAGCTGGATGAG 5815

RESULT 13
US-09-949-016-1242
; Sequence 1242, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIORITY FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1242
; LENGTH: 6861
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-1242

Alignment Scores:
Pred. No.: 7,688-06 Length: 6861
Score: 159.00 Matches: 137
Percent Similarity: 39.15% Conservative: 121
Best Local Similarity: 20.79% Mismatches: 212
Query Match: 4.83% Indels: 192
DB: 4 Gaps: 32

US-10-736-868-2 (1-643) x US-09-949-016-1242 (1-6861)
Qy 15 IleAlaProLeuSerAlaGlnSerProSerThrSerAspAlaProGlyAlaLeuLeuSer 34
Db 4016 GTGGCGTCCCTCAGTTCCAGTCCAGTCCAGGACACCCAGGAGCTG----- 4057
Qy 35 SerLeuValGlyLysSerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGlu 54
Db 4058 ---CTTCAAGAAAGAACCCCGCAGAGAGCTCAACGTGTCTACGAGCTGCGCCAGCTGGAG 4114
Qy 54 ----- 54
Db 4115 GAGGAGCGGAACAGCTGCAAGACAGCAGCTGGACGAGGAGATGAGGCGCAACGAGCCTG 4174
Qy 55 -----LeuMetGlyValGlnPheValAspAlaLeuIleLysGlyGln 69
Db 4175 GAGCGCCACATCTCCACTCTCAACTCCAGTCTCCGACTCGAAG---AAGAAGCTGCGAG 4231
Qy 70 -----MetGluMetAlaLysGlyAlaPheLysThrGlnLeu 81
Db 4232 GACTTTGCCAGCAGCTGGAAGCTCTGGAAGAGGGGAAGAGAGGTTCCAGAGGAGATC 4291
Qy 82 GluValLeuGluLysValHisProAspGln-----PheAspLysTyLysLysLeu 98
Db 4292 GAGAACCTCACCCAGCAGTACGAGGAGAGAGGCGCGCTTATGATTAAGTGGAAAGACC 4351
Qy 99 Lys-----ValAspLeuAlaAlaAspAlaValMetGlnGlnAla 112
Db 4352 AAGAACAGGCTTACGACGAGGCTGACGACCTGGTGTGTGTTGTTGGCAACACGCGGCAA 4411
Qy 113 GluMetAlaLysLeuGlnProLysSerGlyAsnAlaPheIleAspMetLeuAsnGlyAsn 132
Db 4412 CTCGTGTCCAACTGGAAAGAACGAGGAGAA-----TTTGATCATGTTGTTAGCCGAG 4465
```

```
Qy 133 GlyIleProIleGlySerSer----- 139
Db 4466 GAGAAAACATCTCTTCCAAATACCGCATGAGAGGAGCAGAGCTGAGCAGACCCAGG 4525
Qy 140 -----IleArgGlyLeuGluAspAlaIleArgThrGln 150
Db 4526 GAGAAGGAAACCAAGGCCCTCTCCCTGGCTGGGCTTGAAGAGGCTTGAAGCCAAA 4585
Qy 151 ArgAspMetGluAsnThrAspProSerGluGlnIleAlaLysAlaValMetAspLysPhe 170
Db 4586 GAGGAACCTCGAGCGGACCAAC-----AAATGCTCAAGCCGAATGAA----- 4630
Qy 171 GlnThrGlnIleLeuProGlyLeuValAlaAsnMetIleAlaGlyLysAsnProPhe--- 189
Db 4631 -----GACCTGGTTCAGCTCCAGGATCAGCTGGGC 4660
Qy 190 LysMetProGlnGlnMetArgLysAlaGlnAlaProSerSerValPheGlnGlnAla 209
Db 4661 AAGAACGTCCTCATGAGCTGGAGAGTCCAAAGCGGCTTGGAGACCCAGATGGAGGAGATG 4720
Qy 210 LeuAlaGlnArgAlaMetLeuGlyLysAsnAlaProValAlaGlyGlyArgGlyGluGlu 229
Db 4721 AAGACGCGAGCTGGAGAGCTGGAGACGAGCTGCAAGCCACG-----GAGAC 4768
Qy 230 GlnArgMetMetMetAsnArgValAspGlnArgMetGln-----Gln 243
Db 4769 GCCAAACTG-----CGCTGGAAGTCAACATGAGGCGCTCAAGGGCCAGTTCGAA 4819
Qy 244 ArgGluLeuGlnGluAspGluAspAspAspLeuGluAspGluAspValProArg 263
Db 4820 AGGATCTCCAAAGCCCGGACGAGCAGCAATGAG-----GAGAAGAGGAGGCAACTGCAG 4873
Qy 264 ArgArgSerSerAspGlyGluProGlnSerGluAlaGluHisGlnArgArgAspLeuAla 283
Db 4874 AGACAGCTTACAGATGATGACCGAACTGGAAGACGAGCGAAGCAACGCTGCCCTGCA 4933
Qy 284 ArgArgLeuLysSerSerProArgLeuLysGluLeuLeuGlnAsnAlaGluValGlnSer 303
Db 4934 GCTGCAGCAAGAAG-----AAGCTGGAAGGAGCCTGAAAGACCTCGAGCTTCAGGCC 4987
Qy 304 LeuLeuSerTyLysGlnArgMetArgAspSerProLeuSerLysArgArgProLeu----- 321
Db 4988 ---GACTCTGCATCAAGGGGAGGAGGAGCCATCAAGCAGCTACGCAAACTGCAGGCT 5044
Qy 322 AlaMetAsnAsp-----GluAspGluSerAlaPheArg---AlaMetGlu 335
Db 5045 CAGATGAAGGACTTTCAAAGAGAGCTGGAAGATGCCGTGCTCCAGAGATGATGATCTTT 5104
Qy 336 AlaArgAlaLysLeuAspGlnLysSerGlnLeuValLeuGlyLeuHisGlyPheGlyGlu 355
Db 5105 GCCACAGCCAAAGAGATGAGAAAGCCAGAGCTTG-----GAA 5146
Qy 356 SerAspAspAspGluAspGluGluAspGluAsnLeuIleAspProSerGluAsnSerPhe 375
Db 5147 GCAGACCTCATGACGTACAGTACAAAGAGGAC-----CTCGCGCGCTGAGAGGCTCGC 5197
Qy 376 ArgArgAlaProLeu-----ArgLeuSerSerGlyPheValGluLysLeuLysSer 392
Db 5198 AAACAAGCGGACCTCGAAGAGGAGGAACTGGCAGAGAGCTGGCCAGCTAGCTCTCGGGA 5257
Qy 393 AsnAspGluLeuLysSerAlaLeuAspArgIleLysTyLysArgValAspValGluLys 412
Db 5258 AGGAAACCATCTCCAGGACGAGAGAGCGCGCTGGAGCGCGGATCGCCCGCTGGAGGAG 5317
Qy 413 TyLeuAlaProLysProMetGluPheAsnProLysProGlnProGlyTyLysPheAlaPro 432
Db 5318 GAGCTGGAGGAGGAGCAGGCG-CAACATGGAGGCGCATGAGCGACCGCGTCC-GCAAAGCCA 5375
Qy 433 ArgLysIleProThr-----ArgProArgLys 441
Db 5376 CACAGCGCGGAGCAGCTCAGCAACAGCTGGCCACAGAGCGCAGCAGCGCCAGAGA 5435
Qy 442 Met---LeuProLeuLeuIleGlySerAspProLysValGlnGluGluIleArgHis 460
```

```

Db      5436  ATGAGATGCC-----GGCAGCAGCTCGAGCGCGAGAAACAGGAGCTCGGAGCA 5486
Qy      461  ProSerThrGluTrpTyrIleAlaLysGluSerArgValLeuThrAsnLeuLysAsnAsn 480
Db      5487  AGCTCCACGAGATGGAGGG-GCGGTCAAGTCC-----AAGTTCAG 5527
Qy      481  ProSerLeuAlaLeu-----PheMetAspAspLysLeuGluAsnThr 495
Db      5528  TCCACCATCCGCGCTGGAGGCCAAGATTGCACAGCTGGAGGACGAGTTCGAGCAGGAG 5587
Qy      496  LeuLysGlyArgGlnMetLeuThr----- 503
Db      5588  GCCAGAGAAACAGCGCGGCCACCAAGTCGCTGAGCAGAAACAGCAAGAGCTGAAGGAA 5647
Qy      504  -----AspGlu-----Gln 506
Db      5648  ATCTTGCTGCAGGTGGAGGACGAGCGCAAGATGCGCCAGCAGTACAAAGGAGCAGGAGAG 5707
Qy      507  LysGlyArgThrArgValIleThrIleArgAlaLeuProArgLeuPheGlyAlaProThr 526
Db      5708  AAAGGCAATCCAGGGTCAAGCAGCTCAAG-----AGGCAGCTGGAGGAGGCAGAG 5758
Qy      527  AlalysAlaGluMetIleAspAla-----LysValPheGlnAspIleGluGlu 542
Db      5759  GAGGAGTCCCGAGGCATCAAGCCCAACCGCAGGAGAGCTGACGCGGAGCTGGATGAG 5815

```

RESULT 14

```

US-09-902-540-1175/c
; Sequence 1175, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1175
; LENGTH: 19269
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1175

```

```

Alignment Scores:
Pred. No.: 4.46e-05 Length: 19269
Score: 158.50 Matches: 125
Percent Similarity: 37.32% Conservative: 106
Best Local Similarity: 20.19% Mismatches: 215
Query Match: 4.82% Indels: 173
DB: 4 Gaps: 29

```

US-10-736-868-2 (1-643) x US-09-902-540-1175 (1-19269)

```

Qy      27  AspAlaProGlyAlaLeuLeuSerSerLeuValGlyLysSerHisGlnLysLeuProLeu 46
Db      12884  GACGACCCAGTACGCTCGTGGATGAGTGAGCGGGGAGTCGGTTCGAGCGG-----CTC 12831
Qy      47  AlaProSerMetGluAlaLeuGluMetGlyValGlnPheVal----- 61
Db      12830  GCTCAATCGCAGGAGGCGGATGCGTGTCTCGAGGCTTCCTTGTCTCCAGCGTCCCGGC 12771
Qy      62  AspAlaLeuLysLysGlyGlnMetGluMetAlalysGlyAlaPheLysThrGlnLeu 81
Db      12770  AACCGCGTGACGAAGGTGGCACTGGGGGCTTGCTCGTGTGCTCGTGAATCAGTCG 12711
Qy      82  GluValLeuGluLysValHisProAspGlnPheAspLysTyrLysLysLeuLysValAsp 101

```

```

Db      12710  GAA-----GAGCGGTACGACCTGCGCGCTCGATACACCGCTCAAGGCCCGCTCCTT 12657
Qy      102  AspLeuAlaAlaAspAlaValMetGlnGlnAlaGluMetAlalysLeuGln----- 118
Db      12656  GAGAGGTGGAGTCCGCCCTTGAGCAACGAGGAACAGTCCCTCTGAAGGCAATGGAG 12597
Qy      119  -----ProLysSerGlyAsnAlaPheIleAspMetLeuAsn 130
Db      12596  GGGGCTACGCCAGTCTTTCACGCCCCALAGCGGAAAG-----CTCGTCACC 12549
Qy      131  GlyAsnGlyIlePro---IleGlySerSerIleArgGlyLeu---GluAspAlaIleArg 148
Db      12548  GGAAGCGACTCGATCGTGGGTGCTCGCATCGCGGAGTTGCGGAGGAGGAGCTCGG 12489
Qy      149  ---ThrGlnArgAspMetGluAsnThrAspProSerGluGlnIleAlalysAlaValMet 167
Db      12488  TTGACTCAGCGGTGCGGAG-----ATCGAGCAGCTTCCACGAGAGCATG 12441
Qy      168  AspLysPheGlnThrGlnIleLeuProGlyLeuValAlaAsnMetIleAlaGlyLysAsn 187
Db      12440  GACCTCGGCAGAGCAGCTC----- 12420
Qy      188  ProPheLysMetProGlnGlnMetArgLysAlaGlnAlaProSerSerValPheGln 207
Db      12419  ---TCCCTGCGCGAGGAACGAGGACCTATGACGAGCGGCTCGCTAGCTTG---CAG 12369
Qy      208  GlnAlaLeuAlaGlnArgAlaMetLeuGlyLysAsnAlaProValAlaGlyArgGly 227
Db      12368  GAGCGCATCGCGAGGAGGAGAGCTGGAGCAGCGCGCTCCCGGGAAGAGGAGCTC 12309
Qy      228  GluGluGlnArgMetMetMetAsnArgValAspGlnArg-----Met 241
Db      12308  GAGCGTCAGCGAAAGAGTGGCAGGCGCTGGACCAAGCAGCAGCAGAGCTTCTCGGCTC 12249
Qy      242  GlnGlnArgGluLeuGlnGluAsp-----GluAspAspAspLeuLeuAsp 258
Db      12248  CACCAAGAGTGTCTCCAGCAGGAGGACCTGCTGGCTCGCAACGAGCGGCTTTCAGGAG 12189
Qy      259  GluAspValProArgArgArgSerAspGlyGluProGlnSerGluAlaGluHisGln 278
Db      12188  GCCCAGGAGCTCGTCCGCGCGCGCGGAGACGCGGCTCGGGAAGCAGCGCAGCACAAG 12129
Qy      279  ArgArgAspLeuAlaArgArgLeuLysSerSerProArgLeuLysGluLeuGlnAsn 298
Db      12128  -----GCTCAGCAG 12120
Qy      299  AlaGluValGlnSerLeuSerTyrGlnArgMetArgAspSerProLeuSerLysArg 318
Db      12119  GGGGAGTGGAG-----CTGGCGGAGCAGCGCGTGAACGAGGAGCAGCTCTCTCGAA--- 12069
Qy      319  ArgProLeuAlaMetAsnAspGluAspGluSerAlaPheArgAlaMetGluAlaArgAla 338
Db      12068  -----GATGCGCGCT 12057
Qy      339  LysLeuAspGlnLysSerGlnLeuValLeuGlyLeuHisGlyPheGlyGluSerAspAsp 358
Db      12056  TCCCTGGCGGAGCAGCGGAAGTTG----- 12033
Qy      359  AspGluAspGluGluAspGluAsnLeuIleAspProSerGluAsnSerPheArgArgAla 378
Db      12032  GACGGGCTCGTCAAGCAGGCGGCGCGGTGGAGACCGCAGTGGAGAATGCCCCGCAAG 11973
Qy      379  ProLeuArgLysSerSerGlyPheValGluLysLysSerAsnAspGlu----- 395
Db      11972  ACTGCGCGCTGAAGCGCTTCAGACGCGAGAGGTAAAGCGCCCGAGGAGGTGCAGCGC 11913
Qy      396  ---LeuLysSerAlaLeuAspArgIleLysTyrArgValAspAspValGluLysTyrLeu 414
Db      11912  AAGCTGAATCAGCCCTGGACAGCTGGAGCAGCAGCGGGGTTCAGGTC---TCTTCACT 11856
Qy      415  AlaProLysProMetGluPheAsnProLysProGlnProGlyTyrPheAlaProArgLys 434
Db      11855  GCCGAGAGCCCGCAGCGTATCGAGTGGAGGCCAG---GGTCACACACACCGCAAG 11799

```

Qy	435	IleProThrArgProArgLysMetLeuProLeuLeu-----	446
Db	11798	CTCAAGGGGAGGACGAAGAGTCTTCTCGGGGTGCGTGGTGCACTGCGCATCAAG	11739
Qy	447	-----IleGlySerAspProLysVal	453
Db	11738	GGAGTGGTGAAGTCCGCGTGCACATGCGGCCGAGGAGATTGGC-----AAGCTC	11688
Qy	454	GlnGluGluIleArgArgHisProSer-----	462
Db	11687	GAAGCGAGGTTCGAGAAGACACCGGAAGGACCTGGCGCGCGCTCCATGAGCAGCGAGTC	11628
Qy	463	-----ThrGluTtpLysIleAlaLysGluSerArgValLeuThr	475
Db	11627	GATCATGTGTGGTCTGCGCGCGAGTGG-----GCGTCCAGCAGGTCTGTGCTTCAG	11574
Qy	476	AsnLeuLysAsnAsnProSerLeuAlaLeuPheMetAspAspLysLeuGluAsnThr	495
Db	11573	GAGCTGTCGAAGTACGAGGAGGCACGAGCCAGTCTTCGGAGGAC-----AACAAA	11523
Qy	496	LeuLysGlyArgGlnMetLeuThrAspGluGlnLysGlyArgThrArgValLysThrIle	515
Db	11522	GTCGAGAGCTCAGGTGCTCAAGAACACGACGCTGACGACGACGCG-----AAGACG---	11469
Qy	516	ArgAlaLeuProArgLeuPheGlyAlaProThrAlaLysAlaGluMetIleAspAlaLys	535
Db	11468	-----GGGGCACTGAGCGCCCACTCACTTCGCGG-----	11436
Qy	536	ValPheGlnAspIleGluArgProIlePro-----ProLeuPhePheGluPro	552
Db	11435	-----GAGGAGCTACAGACCCACCATTTCCCGAGCTGAAGGCGCTGACGGAGGAATC	11382
Qy	553	LysGlyArgHisThrArgLeuArgTtpThrGlyAlaAsnGluLysGluIleProGly	571
Db	11381	AAGCCCGCAGCAGGAGTGAAGCCCGGGACAAGGCGCGGAGGAGGTGAAGGA	11325

RESULT 15
 US-08-875-435B-1
 ; Sequence 1, Application US/08875435B
 ; Patent No. 6593304
 ; GENERAL INFORMATION:
 ; APPLICANT: Hasegawa, Kazuhide
 ; APPLICANT: Arakawa, Emi
 ; APPLICANT: Oda, Shoji
 ; APPLICANT: Matsuda, Yuzuru
 ; APPLICANT: Takahashi, Katsuhito
 ; APPLICANT: Sugahara, Michihiro
 ; APPLICANT: Ishiyama, Haruo
 ; TITLE OF INVENTION: RECOMBINANT DNA COMPRISING DNA CODING
 ; TITLE OF INVENTION: FOR MYOSIN HEAVY CHAIN SMI ISOPFORM PROTEIN INSERTED INTO
 ; TITLE OF INVENTION: VECTOR DNA, MICROORGANISM CARRYING THE RECOMBINANT DNA, AND
 ; TITLE OF INVENTION: AN AGENT FOR TREATMENT OF ARTERIOSCLEROSIS COMPRISING THE
 ; TITLE OF INVENTION: RECOMBINANT DNA

```
,
, INDEPENDENCE: 07896-0130U.
, CURRENT APPLICATION NUMBER: US/08/875,435B
, PRIOR FILING DATE: 1997-07-25
, PRIOR APPLICATION NUMBER: PCT/JP96/00134
, PRIOR FILING DATE: 1996-01-25
, NUMBER OF SEQ ID NOS: 5
, SOFTWARE: FastSEQ for Windows Version 4.0
, SEQ ID NO 1
```

```

; LENGTH: 6175
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (105)...(6020)
US-08-875-435B-1

```

Alignment Scores:
 Pred. No.: 1.06e-05
 Score: 157.00
 Length: 6175
 Matches: 127

Percent Similarity:	38.31%	Conservative:	127
Best Local Similarity:	19.16%	Mismatches:	211
Query Match:	4.77%	Indels:	198
DB:	4	Gaps:	32

US-10-7336-868-2 (1-643) x US-08-875-435B-1 (1-6175)

Qy	15	IleAlaProLeuSerAlaGlnSerProSerThrSerAspAlaProGlyAlaLeuLeuSer	34
Db	4032	GTGGCTTCCCTTGGATCCCACTTCAGGACACCAAGAGCTG	4073
Qy	35	SerLeuValGlyIysSerHisGlnIysLeuProLeuAlaProSerMetGluAlaLeuGlu	54
Db	4074	---CTCCAGAAGAAACCCGGCAGAAGCTCAATGTGTCTTACCAAGCTGCCTCAGTTGGAA	4130
Qy	54	-----	54
Db	4131	GATCAAGGAACAGCCTGCAGGACCAGCTGGATGAGGAGATGAGGCTAAGCAAAACCTG	4190
Qy	55	-----LeuMetGlyValGlnPheValAspAlaLeuIleIysLysGlyGln	69
Db	4191	GAGCGCATGTCTCAACACTGAACACTTCAGCTCTCAGACTCTAAG---AAGAAGCTGCAG	4247
Qy	70	-----MetGluMetAlaIysGlyAlaPheIysThrGlnLeu	81
Db	4248	GACTTTGCAAGTACCATCGAGTCTATGGAGAGGGGAAGAAGAGGTTTACAGAAGAAGATG	4307
Qy	82	GluValLeuGluIysValHisProAspGln-----PheAspIysValIysLysLeu	98
Db	4308	GAGGGCTCAGCAGCAGTATGAGGAGAGGGCGGCTGCTATGACAACTGGAGAAAACC	4367
Qy	99	Lys-----ValAspAspLeuAlaAlaAspAlaValMetGlnGlnAla	112
Db	4368	AAGAACAGGCTCCAGCAGGAGCTGGATGACTTGGTCTGGACTTGGACAACACCGCGCAA	4427
Qy	113	GluMetAlaIysLeuGlnProIysSerGlyAsnAlaPheIleAspMetLeuAsnGlyAsn	132
Db	4428	CTGGTATCCAAATCTGAAAGAAGACAGAGA---TTTGACCAAGTTGTTAGTGCAG	4481
Qy	133	GlyIleProIleGlySerSer-----	139
Db	4482	GAGAAGACATCTCTCCAAATGTCGGATGAGAGACGCGAGCTGAAGCAGAGGCCAGG	4541
Qy	140	-----IleArgGlyLeuAlaAspAlaIleArgThrGln	150
Db	4542	GAAGAAGGAGACAAAGCTTTGTCTCTAGCCGGGCGCTGGAGGAAGCCCTGGGAAGCAAA	4601
Qy	151	ArgAspMetGluAsnThrAspProSerGluGlnIleAlaIysAlaValMetAspLysPhe	170
Db	4602	GAAGAGCTGGAGGAGCAAC-----AGATGCTCAAGCTGAGATGGAA-----	4646
Qy	171	GlnThrGlnIleLeuProGlyLeuValAlaAsnMet---IleAlaGlyLysAsnProPhe	189
Db	4647	-----GACCTGGTCACTCCAGGATGATGTAGGCAAGAACGTGCAT	4688
Qy	190	LysMetProGlnGlnMetArg-----LysAlaGlnAla	200
Db	4689	GAATGGAGAAGTCCAAGCGTCTTGGAGACCCAGATGGAAGAGATGAAACCCAGCTG	4748
Qy	201	AlaProSerSerValPheGlnGlnAlaLeuAlaGlnArgAlaMetLeuGlyIysAsnAla	220
Db	4749	GAGGAGCTCGGAGGATGACGTGCAGGCC---ACTGAGGATGCCAAGCTCGCGGTAGAGGTC	4805
Qy	221	ProValAlaGlyGlyArgGlyGluGluGlnArgMetMetMetAsnArgValAspGlnArg	240
Db	4806	AACATCAGGCCCTCAAGGGCCAGTTTGAACCGCATCTCCAGCTCCGGATGAACAGAAT	4865
Qy	241	MetGln---GlnArgGluLeuGlnGlu---AspGluAspAspLeuGluAsp	258
Db	4866	GAGGAGAAGAGGCGAGCTACAGCGGAGCTGCAGGATATGACAGACAACTGGGAAT	4925
Qy	259	GluAspValProArgArgArgSerAspGlyGluProGlnSerGluAlaGluHisGln	278

Search completed: July 18, 2005, 16:15:46
Job time : 448 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 18, 2005, 12:25:31 ; Search time 2544 Seconds
(without alignments)
1603.522 Million cell updates/sec

Title: US-10-736-868-2
Perfect score: 3291
Sequence: 1 MILFLFLFLGFCIAPLSA.....SKTRFVGCGAFDMPALGL 643

Scoring table:

BLOSUM62	
Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5
Fgapop 6.0	Fgapext 7.0
Delop 6.0	Delext 7.0

Searched: 7173243 seqs, 3172129809 residues

Total number of hits satisfying chosen parameters: 14346486

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV=xlh
-MODEL=frame+p2n.model
-Q=/cgn2_1/USPTO.spool/US10736868/runat_14072005_105511_2334/app.query.fasta_1.839
-DB=PublishedApplications_NA -QPM=fascp -SUFFX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10736868 @CNG 1.1.480 @runat_14072005_105511_2334
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
- 18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
- 19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq.*
- 20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq.*
- 21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq.*
- 22: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq.*
- 23: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
- 24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
- 25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3291	100.0	1932	21	US-10-736-868-1
2	176.5	5.4	7396	17	US-10-028-248A-35
3	176.5	5.4	7396	17	US-10-107-782-35
4	173.5	5.3	7274	21	US-10-636-909A-48
5	169.5	5.2	2791	20	US-10-425-115-63082
6	169.5	5.2	6126	17	US-10-188-248-25
7	169.5	5.2	6151	18	US-10-302-172-528
8	165	5.0	3603	17	US-10-369-493-27636
9	163.5	5.0	2932	19	US-10-747-065-1
10	163.5	5.0	6354	15	US-10-084-817-158
11	163	5.0	4041	17	US-10-264-049-543
12	162.5	4.9	3257	17	US-10-369-493-30630
13	162.5	4.9	5883	21	US-10-956-157-859
14	162	4.9	3320	19	US-10-283-975A-279
15	159	4.8	3388	9	US-09-954-456-1602
16	159	4.8	3388	9	US-09-967-768A-245
17	159	4.8	3388	9	US-09-954-531-988
18	159	4.8	3388	9	US-09-873-367C-85
19	159	4.8	3388	10	US-10-240-425-1200
20	159	4.8	3388	18	US-10-843-641A-85
21	159	4.8	3388	21	US-10-843-641A-2055
22	159	4.8	3388	21	US-10-843-641A-2449
23	159	4.8	3388	21	US-10-843-641A-4629
24	159	4.8	3388	21	US-10-843-641A-6390
25	159	4.8	3388	21	US-10-887-553A-800
26	159	4.8	4935	21	US-10-956-157-1980
27	159	4.8	5835	21	US-09-927-597-1
28	159	4.8	5835	22	US-10-486-057-1
29	159	4.8	5937	10	US-09-927-597-3
30	159	4.8	5937	22	US-10-486-057-3
31	159	4.8	6861	14	US-10-171-311-161
32	159	4.8	6861	17	US-10-341-434-102
33	159	4.8	6861	22	US-10-923-035-18
34	159	4.8	6900	14	US-10-171-311-163
35	159	4.8	6900	22	US-10-764-425-13
36	159	4.8	6900	14	US-10-116-802-14
37	159	4.8	11065	14	US-10-116-802-14
38	159	4.8	2836	19	US-10-437-963-1292
39	158	4.8	2836	16	US-10-316-253-43
40	157.5	4.8	2348	17	US-10-282-122A-31788
41	157.5	4.8	3486	17	US-10-282-122A-31788
42	157	4.8	6401	15	US-10-171-581-352
43	157	4.8	6401	15	US-10-177-293-314
44	157	4.8	6856	22	US-10-765-700-42
45	157	4.8	7677	10	US-09-814-353-21993

ALIGNMENTS

RESULT 1
US-10-736-868-1
; Sequence 1, Application US/10736868
; Publication No. US20050079160A1
; GENERAL INFORMATION:
; APPLICANT: Solomon, Aaron
; APPLICANT: Morimoto, Richard
; APPLICANT: Betzel, Greg
; TITLE OF INVENTION: OSG-1 Nucleic Acids and Proteins
; FILE REFERENCE: NMESTERN-08451
; CURRENT APPLICATION NUMBER: US/10736,868
; CURRENT FILING DATE: 2003-12-16
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 1932

RESULT 2

US-10-028-248A-35
; Sequence 35, Application US/10028248A
; Publication No. US20030235882A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Patturajan, Meera
; APPLICANT: Vernet, Corine
; APPLICANT: Casman, Stacie
; APPLICANT: Malyankar, Uriel
; APPLICANT: Shenoy, Suresh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gangolli, Esha
; APPLICANT: Miller, Charles
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Taupier Jr., Raymond J
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Smithson, Glennda
; APPLICANT: Zethusen, Bryan
; APPLICANT: Liu, Xiaohong
; APPLICANT: Colman, Steven
; APPLICANT: Tchernev, Velizar
; APPLICANT: Si, Jingsheng
; APPLICANT: Edinger, Shlomit
; APPLICANT: Stone, David
; APPLICANT: Sciore, Paul
; APPLICANT: Millet, Isabelle
; APPLICANT: Rothenberg, Mark
; TITLE OF INVENTION: NO. US20030235882A1el Nucleic Acids and Polypeptides and Methods
; TITLE OF INVENTION: Thereof
; FILE REFERENCE: 21402-222
; CURRENT APPLICATION NUMBER: US/10/028,248A
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/256619
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/262959
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/272408
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/285189
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/308039
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: 60/311266
; PRIOR FILING DATE: 2001-08-09
; NUMBER OF SEQ ID NOS: 211
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
; LENGTH: 7396
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-028-248A-35

Alignment Scores:
Pred. No.: 2,59e-06 Length: 7396
Score: 176.50 Matches: 170
Percent Similarity: 31.25% Conservative: 95
Best Local Similarity: 20.05% Mismatches: 256
Query Match: 5.36% Indels: 329
DB: 17 Gaps: 34

US-10-736-868-2 (1-643) x US-10-028-248A-35 (1-7396)

Qy 12 GlyPheCysIleAlaProLeuSer-----AlaGlnSerProSerThrSerAsp----- 27
Db 3999 GGCTTCTCAGCCAGTCCGACAGCAAGTCCAGCAAGCTCACCAGGACTTCTCCGCGCTGG 4058
Qy 28 AlaProGly-----AlaLeuLeuSerSerLeuValGlyLysSerHisGlnLysLeuPro 45
Db 4059 AGTCCCAGCTCAGACACACTCAGGAGCTG-CTGACGAGGAGGAGAACCGGACAGAGCTGAGC 4117
Qy 46 LeuAlaProSerMetGluAlaLeuGlu----- 54

Db 4118 CTGAGCACCAAGCTCAAGCAGGTGGAGGACGAGAGAAATTCTTCCGGAGCAGCTGGAG 4177
Qy 55 -----LeuMetGlyValGlnPheVal 61
Db 4178 GAGGAGGAGGCCAAGCACCACTGGAGAAAGCAGATGCCACCTCCATGCCCGCTGGCC 4237
Qy 62 AspAlaLeuIleLysLysGlyGlnMetGluMetAlaLysGlyAlaPheLysThrGlnLeu 81
Db 4238 GAC-----ATGAAAAGAGATGGAGCAGCTGTGGGTGCTTGGAAACTCTCTGAG 4288
Qy 82 GluValLeuGluLysValHisProAsp----- 90
Db 4289 GAGGTGAAGAGGAAGCTCCAGAGGACCTGGAGGCGCTGAGCCAGCGCGCAGAGGAGAAG 4348
Qy 91 -----GlnPheAspLysTyLysLysLeuLys-----ValAspAsp 102
Db 4349 GTGGCGCGCTACGACAAGCTGGAGAAGACCAGACGGCGGTGCAGCAGAGAGCTGGACGAC 4408
Qy 103 LeuAlaAlaAspAlaValMetGlnGlnAlaGluMetAlaLysLeuGlnProLysSerGly 122
Db 4409 CTGCTGTGGACCTGGACCACGCGCAGAGCGGTGCACCTGGAGAGAGAGCAGAAG 4468
Qy 123 AsnAlaPheIleAspMetLeuAsnGlyAsnGlyIleProIleGly----- 137
Db 4469 AAG-----TTTGACCAAGCTCTGGCGGAGGAGAACCATCTCTGCCAAGTATGCAGAG 4522
Qy 138 -----SerSerIle 140
Db 4523 GAGCGCACCGGGCTGAGCGGAGCGCCGAGAGGAGACCAAGGCTCTCTCGCTGGCC 4582
Qy 141 ArgGlyLeuGluAspAlaIleArgThrGlnArgAspMetGluAsnThrAspProSerGlu 160
Db 4583 CGGGCCCTGGAGGAAGCATGGACAGAGCGGAGCTGGAGCGGCTCAAC----- 4633
Qy 161 GlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeuProGlyLeuValAla 180
Db 4634 -----AAGCAGTTCCGACGGAG-----ATGAG 4657
Qy 181 AsnMetIleAlaGlyLysAsnProPhe---LysMetProGlnGlnMetArgLysAlaGln 199
Db 4658 GACCTTATGAGCTCAAGGATGATGTGGGCAAGAGTGTCCACGAGCTGGAGAGTC-CAA 4716
Qy 200 AlaAlaProSerSerValPheGlnGlnAlaLeuAlaGlnArgAlaMetLeuLysLysAsn 219
Db 4717 GCGGGCCCT-----AGAGCAGCAGGTGGAGGAGAT 4746
Qy 220 -----AlaProValAlaGlyArgGlyGluGlnArgMetMetMetAsn-Ar 236
Db 4747 GAAGACGACGCTGGAAGAGCTGGAGGACGAGCTGCAGGCCACCGAAGATGCCAAGCTCG 4806
Qy 236 qValAspGlnArgMetGln-----GlnArgGluLeuGlnGluAs 250
Db 4807 GTTGAGGTCAACCTGCAGGCCATGAAGGCCCATGAAGGCCAGTTCGAGCGGACCTTCGAGG 4866
Qy 250 pGluAspAspAspLeuGluAspGluAspValProArgArgArgSerSerAspGlyG 270
Db 4867 CGACGACGCGGAGGAG 4926
Qy 270 uProGlnSerGluAlaGluHisGlnArgArgAspLeuAlaArgLysSerSerPr 290
Db 4927 GCTGGAGGACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4986
Qy 290 oArgLeuLysGluLeu----- 295
Db 4987 GGACCTGAAGGACCTTGAGGCGGCACATCGACTCGGCCCAACAGAACCGGGACGAGCCAT 5046
Qy 296 -----LeuGlnAsnAlaGluValGlnSerLeuSerTyArgGlnArgMetArgAspSe 313
Db 5047 CAACACAGCTCGGAGAGCTGAGGCCCGCAGATGAAGGACTGCATGCCGAGCTGGATGACAC 5106
Qy 313 rProLeuSerLysArgArgProLeuAlaMetAsnAspGluAspGluSerAlaPheArgAl 333


```
; NAME/KEY: CDS
; LOCATION: (140)..(6016)
US-10-107-782-35

Alignment Scores:
Pred. No.: 2,598-06 Length: 7396
Score: 176.50 Matches: 170
Percent Similarity: 31.25% Conservative: 95
Best Local Similarity: 20.05% Mismatches: 256
Query Match: 5.36% Indels: 329
DB: 17 Gaps: 34

US-10-736-868-2 (1-643) x US-10-107-782-35 (1-7396)

QY 12 GlyPheCysIleAlaProLeuSer-----AlaGlnSerProSerThrSerAsp----- 27
Db 3999 GCGTTCTCAGCCAGTCCGACAGCAAGTCCAGCAAGCTCACCAGGACTTCTCCGGCTGG 4058
QY 28 AlaProGly-----AlaLeuLeuSerSerLeuValGlyLysSerHisGlnLysLeuPro 45
Db 4059 AGTCCACAGCTGCAGGACACTCAGGAGCTG-CTGCAGGAGGAGAACCGCGCAGAAAGCTGAGC 4117
QY 46 LeuAlaProSerMetGluAlaLeuGlu----- 54
Db 4118 CTGAGCACCAAGCTCAGCAGGTGGAGGACGAGAAAGATTCCTCCGGGAGCAGCTGGAG 4177
QY 55 -----LeuMetGlyValGlnPheVal 61
Db 4178 GAGGAGGAGGCCAACACACACTGCGAGAGCAGATCGCCACCTCCTCCAGCTGGGCC 4237
QY 62 AspAlaLeuIleLysLysGlyGlnMetGluMetAlaLysGlyAlaPheLysThrGlnLeu 81
Db 4238 GAC-----ATGAAAAGAGATGGAGACAGTGTGGGTGCTCGAAACTGCTGAG 4288
QY 82 GluValLeuGluLysValHisProAsp----- 90
Db 4289 GAGGTGAAGAGGAGACTCCAGAGGACCTGGAGGCGCTGAGCGGCGCAGCAGAGGAGAG 4348
QY 91 -----GlnPheAspLysTyrLysLysLeuLys-----ValAspAsp 102
Db 4349 GTGGCGGCTACGACAGCTGGAGAGACCAAGACGCGGCTGCAGCAGGAGCTGGACGAC 4408
QY 103 LeuAlaAlaAspAlaValMetGlnGlnAlaGluMetAlaLysLeuGlnProLysSerGly 122
Db 4409 CTGCTGTGTGGACCTGAGCACACAGCGCCAGAGCGCGTGTCAACCTGGAGAGAGCAGAG 4468
QY 123 AsnAlaPheIleAspMetLeuAsnGlyAsnGlyIleProIleGly----- 137
Db 4469 AAG-----TTTGACCAAGCTCTGCGGAGGAGAGACCATCTCTGCAAGTATGACAGAG 4522
QY 138 -----SerSerIle 140
Db 4523 GAGCGCAGCGGCTGAGCGGAGGCCCGAGAGAGGAGACCAAGCTCTGCTGCTGGCC 4582
QY 141 ArgGlyLeuGluAspAlaIleArgThrGlnArgAspMetGluAsnThrAspProSerGlu 160
Db 4583 CGGCGCTTGGAGAGCCTAGAGCAGAGGAGCGGAGCTGAGAGCGCTCAAC----- 4633
QY 161 GlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeuProGlyLeuValAla 180
Db 4634 -----AAGCAGTTCGCGACGAG-----ATGGAG 4657
QY 181 AsnMetIleAlaGlyLysAsnProPhe-----LysMetProGlnGlnMetArgLysAlaGln 199
Db 4658 GACCTTATGAGCTCCAAGGATCATGTGGCGAAGAGTGTCCAGAGCTGGAGAGTCC-CAA 4716
QY 200 AlaAlaProSerSerValPheGlnAlaLeuAlaGlnAlaMetLeuGlyLysAsn 219
Db 4717 GCGGCGCCT-----AGAGCAGCAGGCTGGAGGAGAT 4746
QY 220 -----AlaProValAlaGlyIleArgGlyGluGlnArgMetMetAsn-Ar 236
Db 4747 GAAGACGAGCTGGAAGAGCTGGAGGACGAGCTGCAGGCGCCAGCAAGATGCCAAGCTGCG 4806
```

```
QY 236 gValaspGlnArgMetGln-----GlnArgGluLeuGlnGluAs 250
Db 4807 GTTGAGGCTCAACCTGCGAGGCCCATCAAGGCCCATTCGAGCGGACCTGCAGGGCGGGA 4866
QY 250 pGluAspAspAspLeuGluAspGluAspValProArgArgArgSerSerAspGlyG 270
Db 4867 CGAGCAGAGCGGAGGAGAGAAAGAGAGCTGCTCAGACAGGTCGCGGAGATGGAGGAGA 4926
QY 270 uProGlnSerGluAlaGluHisGlnArgArgAspLeuAlaArgLeuLysSerSerPr 290
Db 4927 CTTGAGGAGCAGAGGAGAGCAGCGCTCGATGCGAGTGGCGCGCGGAGAAAGCTGGAGAT 4986
QY 290 oArgLeuLysGluLeu----- 295
Db 4987 GGACTTGAAGACCTGGAGGCGCAGATCGACTCGGCCCAACAAGACCGGAGCAGGCCAT 5046
QY 296 -----LeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrGlnArgMetArgAspSe 313
Db 5047 CAAAACAGCTGCGGAGGCTGCAGGCCAGATGAAGAGACTGCATGCGCAGCTGGATGACAC 5106
QY 313 rProLeuSerLysArgArgProLeuAlaMetAsnAspGluAspGluSerAlaPheArgAl 333
Db 5107 CCGCGCTCTCTGAGGAGATCTTGGCCCGCCAGGCCCAAGAGAAAGAGAGCTGAAGAG 5166
QY 333 aMetGluAla-----ArgAlaLys-- 339
Db 5167 CATGAGGCCGAGATGATCCAGTTGCGAGGAGNACTGCGACCGCGGAGCGTGCACAGCG 5226
QY 339 ----- 339
Db 5227 CCAGGCCCGAGCAGGAGCGGATGAGTGGCTGACGAGATCGCCAAACAGCAGCGGCAAGG 5286
QY 340 -----LeuAspGlnLysSerGlnLeuValLeuGlyLeuHisGlyPheGlyGluSe 356
Db 5287 AGCCCTGCGTGTAGAGGAGAGCGCGTGTGGAGCGCGCATCGCCAGCTGGAGGAGGA 5346
QY 356 rAspAspGluAspGluGluAspGluAsnLeuIleAspProSerProSerGlnAsnSerPheAr 376
Db 5347 GCTGAGGAGGAGCAGGCGCAACACGAGGCTGATCAACGCGCGGTGAAGAGGCCAAGCT 5406
QY 376 gArg-----AlaProLeuArgLysSerSerGlyPheValGluLysLeuLys 391
Db 5407 GCAGCTGACAGATCAACCGGACCTGAACTGGAGCGCGGCGCAGCGCCAGAG----- 5461
QY 391 sSerAsnAspGluLeuLysSerAlaLeuAspArg-----lleLysTyrArgVa 407
Db 5462 ---AACGAGATGCTCGGCAGCAGCTGGAGACGCCAGAACAGGAGCTTAAGGTCAAGCT 5517
QY 407 lAspAspValGlu-----LysTyrLeuAla--ProLysPro----- 418
Db 5518 GCAGGAGATGGAGGCGCACTGTCAAGTCCAGTACAGGCTCCATCACCGGCTCGAGGC 5577
QY 419 -----MetGluPheAsnProLysPro-----GlnProG 428
Db 5578 CAAGATTGCAGCTGGAGGAGCAGCTGGAGCAACAGAGACCAAGAGGCGCCAGGCGCTG 5637
QY 428 lyTyr----- 429
Db 5638 CAAACAGGTCGTCGAGCCGAGAGAAAGCTGAAGATGTGCTGTCGAGGTGGATGACGA 5697
QY 430 -----PheAlaProArgLysIleProThrArgProArgLysMetLeuProLeuLeuIleG 448
Db 5698 GCGGAGGAAACCGCAGCAGTACAGGACAGGCGCGCAGACAGGCATCTACCGGCTGAGACA 5757
QY 448 lySerAspProLysValGlnGluIleArgHisProSer----- 462
Db 5758 GCTCAAGCGGAGCTGGAGGAGGCGC---AAGAGGAGGCGCCAGCGGCGCAACGCTCCCG 5814
QY 462 ----- 462
Db 5815 CCGGAAACTGACGCGCAGCTGGAGGAGCGCCACTGAGACGCGCGCATGCCATGAACCGCGA 5874
```

```

Qy 463 -----ThrG 464
Db 5875 AGTCAGCTCCCTAAAGAACAGCTCAGCGCGGGACCTGCCGTGTGCTGCGCCGCCG 5934
Qy 464 luTrpLys-----IleAlaLysGluSerA 472
Db 5935 ANTGGCCGGAAGGCCCGGGGATGGCTCCGACGAGAGGTAGATGGCAAGCGGATGG 5994
Qy 472 rgValLeuThrAenLeuLysAAsnProSerLeuAlaLysPheMetAspLysL 492
Db 5995 GGCTGAGGCCAAACCTGCCAATAAGCCTTCTCTGACAGCTGAGATGGATGGACAGA 6054
Qy 492 euGluAenThrLeuLysGlyArgGlnMetLeuThrAspGluGlnLysGlyArgV 512
Db 6055 CAGACACCACAGCCTCCCTTCCAG-----ACCCGCGCAG-----6089
Qy 512 allysThrIleArgAlaLeuProArgLeuPheGlyAla-----ProThrA 527
Db 6090 -----CAGCCTCTCCCACTTCTTGGACTGTGTGAACATGCTCCTCCTG 6138
Qy 527 laLysAlaGluMetIleAspAlaLysValPheGlnAspIleGluArgProIleProp 547
Db 6139 CCCTC-----CGCCCGTCCCTC-----6156
Qy 547 roLeuPheGluProLysGlyArgHisThrArgLeuArgTTrpThrGlyAlaAsnGluL 567
Db 6157 CATCCGTTCCCT-----6170
Qy 567 ysGluIleProGlyLeuGlySerArgPheIleLeuProSerLeuAspPro-----T 584
Db 6171 -----CAGGTGTGTGGGCACTTGGCTTCTCTGCTGCATCCCTCCAGCTC 6222
Qy 584 hrMetProAlaLeuAsnThrAlaPheSerThrGlnGlyArgAlaArgAspGluTrpAspT 604
Db 6223 CCTCCCTGCTCGAATCTCATACCAAGACAGAGGCGCGCCGCGCAGAGAGCGACC 6282
Qy 604 hrMetPheLysIleProAsn 610
Db 6283 AGCAGGCTCTCTCAGCCCTCT 6302

RESULT 4
US-10-696-909A-48
; Sequence 48, Application US/10696909A
; Publication No. US20050118604A1
; GENERAL INFORMATION:
; APPLICANT: Lorens, James B.
; APPLICANT: Atchison, Robert E.
; APPLICANT: Fietra, Anabella
; APPLICANT: Holland, Sacha
; APPLICANT: Rigel Pharmaceuticals, Inc.
; TITLE OF INVENTION: Modulators of Angiogenesis and Tumorigenesis
; FILE REFERENCE: 021044-005820US
; CURRENT FILING DATE: 2003-10-29
; PRIOR FILING DATE: 2003-10-17
; PRIOR APPLICATION NUMBER: US 60/512,251
; PRIOR FILING DATE: 2003-10-17
; PRIOR APPLICATION NUMBER: US 60/421,989
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 48
; LENGTH: 7274
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: myosin, heavy polypeptide 9, non-muscle (MYH9)
; OTHER INFORMATION: CDNA
US-10-696-909A-48

Alignment Scores:
Pred. No.: 5,03e-06 Length: 7274
Score: 173.50 Matches: 178
Percent Similarity: 30.71% Conservative: 91

```

```

Best Local Similarity: 20.32% Mismatches: 263
Query Match: 5.27% Indels: 346
DB: 21 Gaps: 35

US-10-736-868-2 (1-643) x US-10-696-909A-48 (1-7274)

Qy 12 GlyPheCysIleAlaProLeuSer-----AlaGlnSerProSerThrSerAsp----- 27
Db 3860 GGCTTCTCAGCCAGTCCGACAGCAAGTCAGCAAGCTCACCAGGACTTCTCCGCCCTGG 3919
Qy 28 AlaProGly-----AlaLeuLeuSerSerLeuValGlyLysSerHisGlnLysLeuPro 45
Db 3920 AGTCCCAGCTGCAGGACACTCAGGACCTG-CTGCAGGAGGAGAACCGGACAGACTGAGC 3978
Qy 46 LeuAlaProSerMetGluAlaLeuGlu-----54
Db 3979 CTGAGCAACCAAGCTCAAGCAGGTGGAGGACGAGAGATTCCTCCGGGAGCAGCTGGAG 4038
Qy 55 -----LeuMetGlyValGlnPhe 60
Db 4039 GAGGAGGAGGAGGCCAAGCACCAACCTGGAGAGACAGATGCCACCCTCATGCCCCAGGTG 4098
Qy 61 ValAspAlaLeuIleLysGlyGlnMetGluMetAlaLysGlyAlaPheLysThrGln 80
Db 4099 GCCGAC-----ATGAAAAAGAGATGGAGCACAGTGTGGGTGCTGGAAACTGCT 4149
Qy 81 LeuGluValLeuGluLysValHisProAsp-----90
Db 4150 GAGGAGGTGAAGAGGAGCTCCAGAAAGGACCTGGAGGGCTGTAGCCAGCGGCACGAGGAG 4209
Qy 91 -----GlnPheAspLysTyLysLysLeuLys-----ValAsp 101
Db 4210 AAGTGGCGCCCTTACCACAGCTGGAGAGACACCAAGCGGCTGCAGCAGGAGCTGGAC 4269
Qy 102 AspLeuAlaAlaAspAlaValMetGlnGlnAlaGluMetAlaLysLeuGlnProLysSer 121
Db 4270 GACCTGCTGTGGACCTGGACCACCGCCAGCGCGGTGCAACCTGGAGAGAGAGCAG 4329
Qy 122 GlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIleProIleGly-----137
Db 4330 AAGAAG-----TTTGACCAGCTCTCTGGCGGAGGAGAACCATCTCTGCCAAGTATGCA 4383
Qy 138 -----SerSer 139
Db 4384 GAGGAGCGGACCGGGCTGAGCGGAGGCCCGGAGAGAGAGACCAAGGCTCTGTCTGCTG 4443
Qy 140 IleArgGlyLeuGluAspAlaIleArgThrGlnArgAspMetGluAsnThrAspProSer 159
Db 4444 GCCCGGGCCCTGGAGGAGCATATGGAGCAGAGGCGGAGCTGGAGCGGCTCAAC-----4497
Qy 160 GluGlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeuProGlyLeuVal 179
Db 4498 -----AAGCAGTTCGCCACCGAG-----ATG 4518
Qy 180 AlaAsnMetIleAlaGlyLysAsnProPhe-----LysMetProGlnGlnMetArgLysAla 198
Db 4519 GAGGACCTTATGAGCTCCAGAGTATGTGGGCAAGAGTGTCCACGAGCTGGAGAGTGC- 4577
Qy 199 GlnAlaAlaProSerSerValPheGlnGlnAlaLeuAlaGlnArgAlaMetLeuGlyLys 218
Db 4578 CAAGCGGGCCCT-----AGAGCAGCAGGCTGGAGGA 4607
Qy 219 Asn-----AlaProValAlaGlyArgGlyGluGlnArgMetMetMetAsn 235
Db 4608 GATGAGACGACGCTGGAAGAGCTGGAGGACGAGCTGCAGGCCACCGAAGATGCCAAGCT 4667
Qy 236 -ArgValAspGlnArgMetGln-----GlnArgGluLeuGlnGlu 249
Db 4668 CGCGTTGGAGTCACTCGAGGCCCATGAAGGCCCATGAGGCGGACCTGCAGGGCCG 4727
Qy 249 uAspGluAspAspAspLeuGluAspGluAspValProArgArgSerSerAspG 269
Db 4728 GGACGAGCAGCGGAGGAGAGAAAGCAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 4787

```

Qy	469	ysGluSerArgValLeuThrAsnIleuLysAsnAsnProSerLeuAlaLaLeuPheMeta	489
Db	5850	AGCGGATGGGGCTCAGGCCAAACCTCCGCAATAAGCCTCTTCTCTCCAGCCTGAGATGG	5909
Qy	489	sPAspLysGluAsnThrLeuLysGlyArgGlnMetLeuThrAspGluGlnLygLYA	509
Db	5910	ATGCACAGACACACCAGCCTCCCTTCCCAG-----ACCCGCGAG-----	5953
Qy	509	rgThrArgVallysThrIleArgAlaLeuProArgLeuPheGlyAla-	524
Db	5954	-----CACGCCTCTCCCCACCTTCTTGGAAGTGTGTGAACATGC	5993
Qy	525	--ProThrAlaLysAlaGluMetIleAspAlaLysValPheGlnAspIleGluGluArgP	544
Db	5994	CTCCTCTGGCCTC-----	6011
Qy	544	rolleProProLeuPhePheGluProLysGlyArgHisThrArgLeuArgTrpThrGlyA	564
Db	6012	CCGTCCCCCATCCCGTTTCCCT-----	6034
Qy	564	laAsnGluLysGluIleProGlyLeuGlyserArgPheIleLeuProSerLeuAspPro-	583
Db	6035	-----CCAGGTTGTTGTTGAGGGCATTGGCTTCCTCTGTGCATCCCC	6077
Qy	584	-----ThrMetProAlaLeuAsnThrAlaPheSerThrGlnGly-ArgAlaArgAsp	600
Db	6078	TTCACACTCCCTCCCTCGCTCGAATCTGATACCAAAGACAGAGGGCCCGGGCCAGGCA	6137
Qy	601	GluTrpAspThrMetPheLysIleProAsn-----	610
Db	6138	GAGAGCGACAGCAGAGCTCTCAGCCCTCTTGCCAAAAGCACAAGATGTTGAGGCGA	6197
Qy	611	---AsnTrpAsnProGlyAspGluValGlyPheLysMetAsn	623
Db	6198	GCAGGCGAGCCCGCGGAGGGCCAGAGTTTCTATGAAT	6239
RESULT 5			
US-10-425-115-63082			
; Sequence 63082, Application US/10425115			
; Publication No. US20040214272A1			
; GENERAL INFORMATION:			
; APPLICANT: La Rosa, Thomas J.			
; APPLICANT: Kovalic, David K.			
; APPLICANT: Zhou, Yinua			
; APPLICANT: Cao, Yongwei			
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated wit			
; FILE REFERENCE: 38-21(53222)B			
; CURRENT APPLICATION NUMBER: US/10/425,115			
; CURRENT FILING DATE: 2003-04-28			
; NUMBER OF SEQ ID NOS: 369326			
; SEQ ID NO 63082			
; LENGTH: 2791			
; TYPE: DNA			
; ORGANISM: Zea mays			
; FEATURE:			
; OTHER INFORMATION: Clone ID: MRT4577_157530C.1			
US-10-425-115-63082			
Alignment Scores:			
Pred. No.:	3	44e-06	Length: 2791
Score:	169.50	Matches:	140
Percent Similarity:	37.73%	Conservative:	103
Best Local Similarity:	21.74%	Mismatches:	226
Query Match:	5.15%	Indels:	177
DB:	20	Gaps:	29
US-10-736-868-2 (1-643) x US-10-425-115-63082 (1-2791)			
Qy	20	AlaGlnSerProSerThrSerAspAlaProGlyAlaLeuLeuSerSer	35
Db	76	GCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCAGCCCTTCTTCTTGGCGCGTCATC	135

Db 1706 GTAGATCGCTGAGATTACGACTGGAAGAAAAAGAAATCTTTCTCTCAATATAAAAAACAAA 1765
Qy 498 GlyArgGlnMetLeuThrAspGluGlnLysGlyArgThrArgValLysThrIleArgAla 517
Db 1766 CAGCTACAGGACCTCACA--GAAGAGAGGGG----- 1795
Qy 518 LeuProArgLeuPheGlyAlaProThrAlaLysAlaGluMetIleAspAlaLysValPhe 537
Db 1796 -----ACACTGGCGCGTGAATTCGTGACATGAAGAATATG 1831
Qy 538 GlnAspIleGluArgProIleProProLeuPhePheGluProLysGlyArgHisThr 557
Db 1832 TTAGAAGTGAAGGAAGAAAATCAATGTTCTTCAGAAAAGATTGAAAACCTTGCAAGAA 1891
Qy 558 ArgLeuArgTrpThrGlyAlaAsnGluLysGluIleProGlyLeuGlySerArgPheIle 577
Db 1892 CAACCTTAGG-----GATAAAGACACAGCACTCCCAACCTGAAAGACAGA----- 1936
Qy 578 LeuProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSerThr----- 594
Db 1937 GTGAAGTCCCTTGCAGACGGATTCCAGTAATACAGATACTGCACCTGGCGCAGCTAGAGGAA 1996
Qy 595 -----GlnGlyArgAlaArgAspGlu 601
Db 1997 GCTCTGTGCAGAGGAGAGAAATTAATGAGCGCTTGAAAGACAGCGAGAAAGAGATGAT 2056
Qy 602 TrpAspThrMetPheLysIleProAsnAsnTrpAsnProGlyAspGluValGlyPheLys 621
Db 2057 CGGGAAGACTAGAAAGATAGATAGATCTCCGAAAAGAGAACAAAGACCTGAAAGAGAA 2116
Qy 622 MetAsnSer 624
Db 2117 GTCAATGCT 2125

RESULT 7

US-10-302-172-528
; Sequence 528, Application US/10302172
; Publication No. US20040053250A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20040053250A1el Arginine-rich Protein-like Nucleic Acids an
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 803_ICNCP
; CURRENT APPLICATION NUMBER: US/10/302,172
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/225,251
; PRIOR FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: PCT US02/05095
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 09/799,451
; PRIOR FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 950
; SOFTWARE: PE_FL_genes Version 2.0
; SEQ ID NO 528
; LENGTH: 6151
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (263)..(3133)
US-10-302-172-528

Alignment Scores:

Pred. No.:	1e-05	Length:	6151
Score:	169.50	Matches:	118
Percent Similarity:	36.28%	Conservative:	108
Best Local Similarity:	18.94%	Mismatches:	248
Query Match:	5.15%	Indels:	149
DB:	18	Gaps:	22

US-10-736-868-2 (1-643) x US-10-302-172-528 (1-6151)

Qy 36 LeuValGlyLysSerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGluLeu 55
Db 632 CTGACTGGCTCATCCCATCATCACCACCAGGTCCCTCCCATG---TTGAGGACAGGTA 698
Qy 56 MetGlyValGlnPheValAspAlaLeuIleLysLysGlyGlnMetGluMetAlaLysGly 75
Db 689 AGAGACAGACACATGTTAGATCTTCAGGCCCGCTGAAAGAACTGCAGAGAGAAATGAC 748
Qy 76 AlaPheLysThrGlnLeuGluValLeuGluLysValHisProAspGlnPheAspLysTyr 95
Db 749 CTCCTCCGGAAGAGCTAGACATCAAGGACAGCAAAATGGGATCTTCCATGAACAGATTT 808
Qy 96 LysLysLeuLysValAspAspLeuAlaAspAlaValMetGlnGlnAlaGluMetAla 115
Db 809 AAGACTTCTGGAGTCTCTGAGCTTAAGAGGAGAGAGTCTTTAGGAAAAAGAGGAGCG 868
Qy 116 LysLeuGlnProLysSerGlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIlePro 135
Db 869 CGGATGCTGCTCTCAAGGAGCAGATGAGGGTTTCCATCAAGAAATCAGACCTACAG 928
Qy 136 IleGlySerSerIleArgGlyLeuGluAspAlaIleArgThrGlnArgAspMetGluAsn 155
Db 929 TTG-----ACAATCCAGGCCCTTCAAGATGAGCTGCGAACCCAGAGAGACCTCAACCAC 982
Qy 156 ThrAspProSerGlnGlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeu 175
Db 983 CTCCTCCAGCAGAGAGTGGCAACCGAGGAGCG---GAGCAGCTTCACCATCGAGCTGACC 1039
Qy 176 ProGlyLeuValAlaAsnMetIleAlaGlyLysAsnProPheLysMetProGlnGlnMet 195
Db 1040 -----GAGGAGAACTTTAGCGCGCTCCAAGCCGAG 1069
Qy 196 ArgLysAlaGlnAlaAlaProSerSerValPheGlnGlnAlaLeuAlaGlnArgAlaMet 215
Db 1070 CATGACAGCGCAGCTAAGGAGCTGTTCCTTTTGGAGAGACATTAGAGAAATGGAGCTG 1129
Qy 216 LeuGlyLysAsnAlaProValAlaGlyArgGlyGluGluGlnArgMetMetMetAsn 235
Db 1130 -----AGAAATGAAACGCGAGAAACAAACCTCAAT 1159
Qy 236 ArgValAspGlnArgMetGlnGln-----ArgGluLeuGlnGluAspGluAspAsp 253
Db 1160 GCCCGAGATGAGTCAATTAATAAACTTCTTGAGATGTGCAAAAGTAAAGGCTTGCCATCC 1219
Qy 254 AspAspLeuGluAspGluAspValProArgArgSerSerAspGlyGluProGln--- 272
Db 1220 AAAAGCCTGGAGGATGACAAATGACGAAACCGCGCGATGCGAGGCTGAGTCTCAGGTC 1279
Qy 273 -----SerGluAlaGluHisGlnArgArgAsp 281
Db 1280 AGCCACTTGGAACTGATTTTAGATCAGAAAGAGAGAAACATACATCTTAGAGAGGAA 1339
Qy 282 LeuAlaArgArgLeuLysSerSerProArgLeuLysGluLeuGlnAsnAlaGluVal 301
Db 1340 TTGACCGAAGAGCCAACTTCAGCG-----GAGCCAGCAAGACG 1381
Qy 302 GlnSerLeuLeuSerTyrGlnArgMetArgAspSerProLeuSerLysArgArgProLeu 321
Db 1382 AAGGCTCTCCAGACTGTCTCGAAATGAGGACACAAAAATCGCTTCA-----TTG 1432
Qy 322 AlaMetAsnAspGluAsp---GluSerAlaPheArgAlaMetGluAlaArgAlaLysLeu 340
Db 1433 GAACGAAACATAAGGGATCTTGAGGATGAGATCCAGATGTTAAAGCCAAATGTTGTCTG 1492
Qy 341 AspGlnLysSerGlnLeuValLeuGlyLeuHisGlyPheGlyGluSerAspAspGlu 360
Db 1493 -----AACACTGAG 1501
Qy 361 AspGluGluAspGluAsnLeuIleAspProSerGluAsnSerPheArgArgAlaProLeu 380
Db 1502 GACCGCGAAGAGAG----- 1516

```
QY 381 ArgLeuSerSerGlyPheValGluLysLeuLysSerAsnAspGluLeuLysSerAlaLeu 400
Db 1517 -----ATCAAAACAAATGAGGTTCACAAAGTCACTCC 1549
QY 401 AspArgIleLysTyrArgValAspValGluLysTyrLeuAlaProLysProMetGlu 420
Db 1550 AGTTTATGACCAACAGATTGTCAGCTGAAGAGGAACTTTCAAAGAAAGAGTCGGAA 1609
QY 421 PheAsnProLysProGlnProGlyTyrPheAlaProArgIleProThrArgProArg 440
Db 1610 CTT-----CTTGCTTTACAAACAAAGCTTGAAACCTTCACCAAT 1648
QY 441 LysMetLeuProLeuLeuIleGlySerAspProLysValGlnGluLysLeuArgHis 460
Db 1649 CAA-----AATTCAGATTGCAAGCAACACATTGAAAGTGTCAAAGAG 1690
QY 461 ProSerThrGluTrpLysIleAlaLysGluSerArgValLeuThrAsnLeuLysAsnAsn 480
Db 1691 TCACTTACT-----GCCAAGAACAGAGGCTGCCATC---CTTCAGACTGAG 1735
QY 481 ProSerLeuAlaAlaLeuPheMetAspAspLys-----LeuGluAsnThrLeuLys 497
Db 1736 GTAGATGCGCTGAGATTACGACTGCAAGAAAGAAATCTTCTCAATAAATAACAA 1795
QY 498 GlyArgGlnMetLeuThrAspGlnLysGlyArgThrArgValLysThrIleArgAla 517
Db 1796 CAGCTACAGGACCTCACA---GAAGAGAGGGG-----1825
QY 518 LeuProArgLeuPheGlyAlaProThrAlaLysAlaGluMetIleAspAlaLysValPhe 537
Db 1826 -----ACCTGGCGGTGAATTCGTGACATGAAGATATG 1861
QY 538 GlnAspIleGluGluArgProIleProLeuPhePheGluProLysGlyArgHisThr 557
Db 1862 TTGAAGTGAAGAAAGAAATCAATGTTCTTCAGAAAGAGATGAAACTTCACAGAA 1921
QY 558 ArgLeuArgTrpThrGlyAlaAsnGluLysGluIleProGlyLeuGlySerArgPheIle 577
Db 1922 CAACCTTAGG-----GATAAGACAAAGCAACTGACCAACTGAAAGACAGA-----1966
QY 578 LeuProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSerThr-----594
Db 1967 GTGAAGTCTTCGACAGCGATTCCAGTAATACAGATCTGCACTGGCGACCTAGAGGAA 2026
QY 595 -----GlnGlyArgAlaArgAspGlu 601
Db 2027 GCTCTGTGACAGAGGAGAGATATTGAGCGCTTGAAGAACACAGCGAGAAAGAGATGAT 2086
QY 602 TrpAspThrMetPheLysIleProAsnAsnTrpAsnProGlyAspGluValGlyPheLys 621
Db 2087 CGGGAAGACTAGAAAGAGATAGATCTCTCCGAAAGAGAAACAAAGACCTGAAAGAGAG 2146
QY 622 MetAsnSer 624
Db 2147 GTCAATGCT 2155
```

RESULT 8

```
US-10-369-493-27636
; Sequence 27636, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
```

```
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 27636
; LENGTH: 3603
; TYPE: DNA
; ORGANISM: Neurospora crassa
US-10-369-493-27636
```

Alignment Scores:

```
Pred. No.: 1.36e-05 Length: 3603
Score: 165.00 Matches: 163
Percent Similarity: 33.25% Conservative: 120
Best Local Similarity: 19.15% Mismatches: 259
Query Match: 5.01% Indels: 310
DB: 17 Gaps: 36
```

US-10-736-868-2 (1-643) x US-10-369-493-27636 (1-3603)

```
QY 19 SerAlaGlnSerProSerThrSerAspAlaProGlyAlaLeuLeuSerSer-----35
Db 203 ACGGCATCAAGCCCGTCTTCGTCTTGACGGCGCGCCCGCTCTCAAGCGCGCACCA 262
QY 36 -----LeuValGlyLysSerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeu 53
Db 263 TCCAGGCTCGCGCGCAAGACGAGAGGCGCAGCGATGACGCCACG-CGAAACGCGCAGG 321
QY 54 GluLeuMetGlyValGlnPhe-----Val 61
Db 322 AAATTACTAGCGGTGCAGATCATCTGATGCGAGGAGGAGACGACGAGCGGCGCAAA 381
QY 62 AspAlaLeuLysLysGlyGlnMetGluMetAlaLysGlyAlaPheLysThrGlnLeu 81
Db 382 AGGCAGAGGCGCAGAAAGCAATAAAACGCGCGCAGCAGCAGCAGCAGCAGAA 441
QY 82 GluValLeuGluLysValHisProAspGlnPheAspLysTyrLysLysLeuLysValAsp 101
Db 442 GAA-----CAGGAAGTACTGCGCGCATG-----GACAAAGTCTTTACGCCAC 486
QY 102 AspLeuAlaAlaAspAlaValMetGlnAlaGluMetAlaLysLeuGlnProLysSer 121
Db 487 GAATTAGCATGTCA-----AACACAGACGCCCAAAAGTCTCGACGCTTCACAGCAA 540
QY 122 GlyAsnAlaPheIleAspMetLeuAsnGlyAsn-----GlyIleProIleGly 137
Db 541 GACGCTTACCATCTCTCTGACCTCGACGCGCGCATCGAAGCCATGGGCAAGCGGATGAT 600
QY 138 SerSerIleArgGlyLeuGluAspAlaIleArgThrGlnArgAspMetGluAsnThrAsp 157
Db 601 CCGCGCATCATGTCTCTGGAAGAACTCGAAGAGTACGCTAGGCGATTTGAGAGACGGGAG 660
QY 157 -----157
Db 661 GATATCAACTGTGATGATTTCAGCAAGATCGACTTTGATGGGAGTCTTCAAGTCTTTG 720
QY 158 ---ProSerGluGlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeuPro 176
Db 721 CCGCGCGCTGATCGGTACCAACATCTTGAATCGCGGAGGTTCGAGAGCAGATTCGAAATG 780
QY 177 GlyLeuValAlaAsnMetIleAlaGlyLysAsnProPheLysMetProGlnGlnMetArg 196
Db 781 GGGTTGAGTAAAGACGAGTGTGGACGTGATGTTCCCTAACCGGATGGATTTTCAGCGGGTT 840
QY 197 LysAlaGlnAlaAlaProSerSerValPheGlnGlnAlaLeuAlaGlnArgAlaMet---215
Db 841 CAGATTGAGAGATCAAGGAG-----AGGAATCACTTGACGACGAGACTGATGTATC 891
QY 215 -----215
Db 892 GAGATGGCATGACGGGATTTGGATCTGACGCTGCGCAGTCAACGCGCAGAGAGTGGCGGCG 951
QY 216 -----LeuGlyLysAsnAlaProValAlaGlyGlyArgGly---227
Db 952 GATCAGAAATAGGAGTATATCTCTGTCAGAAATGAAGGTGCGGAAGCGGATATGCGCTG 1011
```


Score: 163.50 Matches: 118
Percent Similarity: 36.17% Conservative: 107
Best Local Similarity: 18.97% Mismatches: 250
Query Match: 4.97% Indels: 147
DB: 19 Gaps: 22

US-10-736-868-2 (1-643) x US-10-747-065-1 (1-2932)

QY 36 LeuValGlySerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGluLeu 55
DB 428 CTGGGTGGCTATCCACCACCACCACCAGGTGCCCTCCATG---TTGAGGAGGTA 484
QY 56 MetGlyValGlnPheValAspAlaLeuLeuLysGlyGlnMetGluMetAlaLysGly 75
DB 485 AGAGACAGCAGCAGTGTGGATCTTCAAGCCAGCTCAAGAACTGCAGAGAGAAATGAC 544
QY 76 AlaPheLysThrGlnLeuGluValLeuGluLysValHisProAspGlnPheAspLysTyr 95
DB 545 CTCCTTCGCAAGAGCTTGACATTAAAGGACAGCAAACTGGGTCTTCCATGATAGCATC 604
QY 96 LysLysLeuLysValAspAspLeuAlaAlaAspAlaValMetGlnGlnAlaGluMetAla 115
DB 605 AAGACTTTCTGGAGTCTCTGAGCTTAAGAGGAGAGAGTCTTGAGGAGAGAGGAGCTGCT 664
QY 116 LysLeuGlnProLysSerGlyAsnAlaPheLeuAspMetLeuAsnGlyAsnGlyLeuPro 135
DB 665 CGCATGTCTGTCTCTCAAGAGCAGATGAGGTTTCTCAGAAAGAAACCCAGCACCTGCAG 724
QY 136 IleGlySerSerIleArgGlyLeuGluAspAlaIleArgThrGlnArgAspMetGluAsn 155
DB 725 TTG-----ACCATCAGGCCCTTCCAGATGAACCTGCGGACCCAGAGAGACTCAACCAC 778
QY 156 ThrAspProSerGluGlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeu 175
DB 779 CTCCTCGCAGCAGAGGTGGCAACCGAGGAGCA---GAGCATTTCCATCGAGCTGACG 835
QY 176 ProGlyLeuValAlaAsnMetIleAlaGlyLysAsnProPheLysMetProGlnGlnMet 195
DB 836 -----GAGGAGAACTTCGCGAGGCTCCAGCCGAA 865
QY 196 ArgLysAlaGlnAlaProSerSerValPheGlnGlnAlaLeuAlaGlnArgAlaMet 215
DB 866 CACGACAGCGAGCGCAAGAGCTGTCTTCGCGAAGAGATTGCAAGAAATGGAGCTA 925
QY 216 LeuGlyLysAsnAlaProValAlaGlyArgGlyGluGlnGluArgMetMetAsn 235
DB 926 -----AGGATTGAACACAGAAACAGACTCTCAAT 955
QY 236 ArgValAspGlnArgMetGlnGln-----ArgGluLeuGlnGluAspGluAspAsp 253
DB 956 GCGCGCAGCAGTCCATTAAGAGCTCTGGAGATGCTGCAGAGTAAGGGCTTGCCATCG 1015
QY 254 AspAspLeuGluAspGluAspValProArgArgArgSerSerAspGlyGluProGln--- 272
DB 1016 AAAAGCCTAGAGGAGCAGCAACAGCGCACACCGCGGATGGCGGCTGAGTCTCAGGTC 1075
QY 273 -----SerGluAlaGluHisGlnArgArgAsp 281
DB 1076 AGCCACTTGGAGTATTTAGACAGAGAGAGAGAAACATCCACTCAGAGAGGAA 1135
QY 282 LeuAlaArgArgLeuLysSerSerProArgLysGlyLeuLeuGlnGlnAsnAlaGluVal 301
DB 1136 TTGACCCGAAGAGCAACTTTCAGCG-----GAGCCAGCCAGAGAGC 1177
QY 302 GlnSerLeuLysSerTyrGlnArgMetArgAspSerProLysSerLysArgArgProLeu 321
DB 1178 AAGCGCTCCAGACTGTCCATCAATGAAGACACAAATAATGTCTCA-----CTG 1228
QY 322 AlaMetAspGluAspGluSerAlaPheArgAlaMetGluAlaArgAlaLysLysAsp 341
DB 1229 GAGCGGAACATCCGGGAC-----CTCGAC 1252
QY 342 GlnLysSerGlnLeuValLeuGlyLeuHisGlyPheGlyGlyGlyGlyGlyGlyGlyGly 361

DB 1253 GATGAGATCCAGATG---TTGAAAGCCCAATGTGTGTCTCAACACAGAGAGCCGAGGAG 1309
QY 362 GluGluAspGluAsnLeuLeuAspProSerGluAsnSerPheArgArgAlaProLeuArg 381
DB 1310 GAGATCAAAACAGATCGAGGTGTACAAAAGCCACTCAAGTTTATGAAGACC----- 1360
QY 382 LeuSerSerGlyPheValGluLysLeuLysSerAsnAspGluLeuLysSerAlaLeuAsp 401
DB 1361 -----AAGATGACCTGAAA----- 1378
QY 402 ArgIleLysTyrArgValAspValGluLysTyrLeuAlaProLysProMetGluPhe 421
DB 1379 -----CAGGAACCTT 1387
QY 422 AsnProLysProGlnProGlyTyrPheAlaProArgLysIleProThrArgProArgLys 441
DB 1388 TCCAGAAGGAGTACAGAACTCTTCGCTTCAAAACAAAGCTTGAACCTCTAGCAATCAG 1447
QY 442 MetLeuProLeuLeuIleGlySerAspProLysValGlnGluGluIleArgArgHisPro 461
DB 1448 -----AATTCAGATTGCAAGCAACACATTTGAAGTCTTAAAGAGTCA 1489
QY 462 SerThrGluTrpLysIleAlaLysGluSerArgValLeuThrAsnLeuLysAsnAsnPro 481
DB 1490 CTACT-----GCCAAAGACAGAGGCTGCCATC---CTTCAGACTCAGGTA 1534
QY 482 SerLeuAlaAlaLeuPheMetAspLys-----LeuGluAsnThrLeuLysGly 498
DB 1535 GATCAGCTGAGATTACCGCTCGAAGAGAAAGAAATCTTTCTCAATAAGAAACAAACAG 1594
QY 499 ArgGlnMetLeuThrAspGluGlnLysGlyArgThrArgValLysThrIleArgAlaLeu 518
DB 1595 CTCCAAGACCTCACT---GAAGAGAGAGGG----- 1621
QY 519 ProArgLeuPheGlyAlaProThrAlaLysAlaGluMetIleAspAlaLysValPheGln 538
DB 1622 -----ACCTAGCTGAGAGATCCGTGATATGAAAGATATGTTA 1660
QY 539 AspIleGluArgProIleProProLeuPheGluProLysGlyArgHisThrArg 558
DB 1661 GAAGTAAAGGAAAGAAATAATCAATGTTCTTCAGAAAAAATTTGAAACCTTCAGAAACAA 1720
QY 559 LeuArgTrpThrGlyAlaAsnGluLysGluIleProGlyLeuGlySerArgPheIleLeu 578
DB 1721 CTTAGG-----GATAAGGACAAACACTGACCACTGAAAGACAGA-----GTG 1765
QY 579 ProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSerThr----- 594
DB 1766 AAGTCCCTGCAGACGGACTCCAGCAACACTGCACACTGCTCTGGCCACTCTCGAGGAGGCC 1825
QY 595 -----GlnGlyArgAlaArgAspGluTrp 602
DB 1826 TTGTCGGAAGAGAGAGAGAAATAAGAGCGCTTGAAGAGCAGAGGAGAGAGATGATCGG 1885
QY 603 AspThrMetPheLysIleProAsnAsnTrpAsnProGlyAspGluValGlyPheLysMet 622
DB 1886 GAAAGACTAGAGAGATAGATCCTTTTCGAAGAGAGAAACAAAGACCTCAAGAGAGAGTGC 1945
QY 623 AsnSer 624
DB 1946 AATGCT 1951

RESULT 10
US-10-084-817-158
; Sequence 158, Application US/10084817
; Publication No. US20030119009A1
; GENERAL INFORMATION:
; APPLICANT: Susan Stuart
; APPLICANT: Jed G. Nuchtern
; APPLICANT: Sharon E. Plon
; APPLICANT: Jason M. Shohet
; TITLE OF INVENTION: GENES REGULATED BY MYCN ACTIVATION


```
QY 430 -----PheAlaProArgLysIleProThrArgProArgLysMetLeuProL 445
Db 5761 GGATGACGAGCGGAGGAAACCGCAGCAGTACAAAGGACCGAGCGCATCTACCGG 5820
QY 445 euLeuIleGlySerAspProLysValGlnGluGluIleArgArgHisProSer 462
Db 5821 CCTGAGCAGCTCAAGCGGCAGCTGGAGAGCGCG---AAGAGGAGGCCCGCGGCCAA 5877
QY 462 ----- 462
Db 5878 CGCTCTCCGCGGAACTGACGCGGAGCTGGAGGAGCGCCACTGAGAGCGCGCATGCCAT 5937
QY 462 ----- 462
Db 5938 GAACCGCGAAGTCAGCTCCCTAAAGACAAGCTCAGCGCGGGGACCTGCCGTTGTGCT 5997
QY 463 -----ThrGluTrpLys 469
Db 5998 GCCCGCGGATGGCGCGGAAAGCGCGGGGATGGCTCCGACGAGAGGTAGATGGCAA 6057
QY 469 ysGluSerArgValLeuThrAsnLeuLysAsnAsnProSerLeuAlaAlaLeuPheMetA 489
Db 6058 AGCGGATGGGCTGAGGCGCAACCTGCCGAATAAGCCTCTCTCTGCGAGCTGATGG 6117
QY 489 spAspLysLeuGluAsnThrLeuLysGlyArgGlnMetLeuThrAspGluGlnLysGlyA 509
Db 6118 ATGGACAGACACACACAGCTCCCTTCCAG-----ACCCCGCAG----- 6161
QY 509 rgThrArgValLysThrIleArgAlaLeuProArgLeuPheGlyAla 524
Db 6162 -----CAGCGCTCTCCCACTTCTTGGGACTCTGTGAACATGC 6201
QY 525 ---ProThrAlaLysAlaGluMetIleAspAlaLysValPheGlnAspIleGluGluArgp 544
Db 6202 CTCCTCTGCGCTC-----CGCC 6219
QY 544 rolleProProLeuPhePheGluProLysGlyArgHisThrArgLeuArgTrpThrGlyA 564
Db 6220 CGTCCCGCCATCCCGCTTTCCT----- 6242
QY 564 laAsnGluLysGluIleProGlyLeuGlySerArgPheIleuProSerLeuAspProT 584
Db 6243 -----CAGGTGTGTGGAGGCATTTGGCTTCTCTGCTGATATCCC- 6284
QY 584 hrMetProAlaLeuAsnThrAlaPheSerThrGlnGlyArgAlaArgAspGluTrpAspT 604
Db 6285 --CTTCCAGCTCCCTCCCTGCTCAGATCTG----- 6314
QY 604 hrMetPheLysIleProAsnAsnTrpAsnProGly 615
Db 6315 -----ATACCAAGAGACAGGCGCGCGGT 6338
```

RESULT 11

```
US-10-264-049-543
; Sequence 543, Application US/10264049
; Publication No. US2004000579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P4133P1
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; SOFTWARE: Patent In Ver. 3.1
; SEQ ID NO 543
; LENGTH: 4041
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```

```
; NAME/KEY: misc feature
; LOCATION: (162)..(162)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4022)..(4022)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4039)..(4039)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-264-049-543
```

```
Alignment Scores: 2.51e-05 Length: 4041
Score: 163.00 Matches: 177
Percent Similarity: 30.23% Conservative: 89
Best Local Similarity: 20.11% Mismatches: 262
Query Match: 4.95% Indels: 354
DB: 17 Gaps: 35
```

```
US-10-736-868-2 (1-643) x US-10-264-049-543 (1-4041)
QY 12 GlyPheCysIleAlaProLeuSer-----AlaGlnSerProSerThrSerAsp----- 27
Db 582 GGCTTCTCAGCCAGTCCGACAGCAAGTCCAGCAAGCTCACCAGGACTTCTCCGGCTGG 641
QY 28 AlaProGly-----AlaLeuLeuSerSerLeuValGlyLysSerHisGlnLysLeuPro 45
Db 642 AGTCCAGCTCCAGGACACTCAGGAGCTG-CTGCAGGAGGAGAACCCGCGAAGCTGAGC 700
QY 46 LeuAlaProSerMetGluAlaLeuGlu----- 54
Db 701 CTGAGCACCAGCTCAGCAGCTGGAGGAGCAGAGAAATTCCTTCCGGGAGCAGCTGGAG 760
QY 55 -----LeuMetGlyValGlnPhe 60
Db 761 GAGGAGGAGGAGGAGCAACACACCTGTGAGAAGCAGATCGCCACCTCCATGCCAGGTG 820
QY 61 ValAspAlaLeuIleLysGlyGlnMetGluMetAlaLysGlyAlaPheLysThrGln 80
Db 821 GCCGAC-----ATGAAAAGAGAGATGAGGACAGTGTGGGTGCTCGAAACTGCT 871
QY 81 LeuGluValLeuGluLysValHisProAsp----- 90
Db 872 GAGGAGGTGAGAGGAGAGCTCCAGAGGACCTCGGGGGCCTGAGCCAGCGCGCAGGAG 931
QY 91 -----GlnPheAspLysTyLysLysLeuLys-----ValAsp 101
Db 932 AAGGTGGCGCGCTACGACAGCTGGAGAAGACCAAGACGCGGCTGCAGCAGAGCTGGAC 991
QY 102 AspLeuAlaAlaAspAlaValMetGlnGlnAlaGluMetAlaLysLeuGlnProLysSer 121
Db 992 GACCTGTGTGGACCTGGACACCAGCGCGCAGCGGTGCAACCTGTGAGAGAAGACAG 1051
QY 122 GlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIleProIleGlySerSerIleArg 141
Db 1052 ---NAGAGTTTGACCACTCTG-----GCCGAGGAGAG 1084
QY 142 GlyLeuGluAspAlaIleArgThrGlnArgAspMetGluAsnThrAspProSerGluGln 161
Db 1085 ACCATCTCTCCAAGTATGACAGAGAGCGCGAGCCGGGTGAGCGGCGGCGGAGAGAAG 1144
QY 162 IleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeu-ProGlyLeuValAlaAs 181
Db 1145 GAGACCAAGGCTCTGTCG-----CTGGCCCGGG----- 1172
QY 181 nMetIleAlaGlyLysAsnProPheLysMetProGlnMetArgLysAlaGlnAlaA1 201
Db 1173 -----CCCTGGAGGAGCCCTGAGCAGAGAGCGGAGCTGGAGCGGC 1213
QY 201 aProSerSerValPheGlnGlnAlaLeuAlaGln-----Ar 213
```


TYPE: DNA
ORGANISM: Caenorhabditis elegans
US-10-369-493-30630

Alignment Scores:
Pred. No.: 2,1e-05 Length: 3257
Score: 162.50 Matches: 132
Percent Similarity: 37.14% Conservative: 99
Best Local Similarity: 21.22% Mismatches: 242
Query Match: 4.94% Indels: 151
DB: 17 Gaps: 28

US-10-736-868-2 (1-643) x US-10-369-493-30630 (1-3257)

```

Qy 18 LeuSerAlaGlnSerProSerThrSerAspAlaProGlyAlaLeuLeuSerSerLeuVal 37
Db 859 ATCGAAGCTCAATATGAGAGCTTCAAGAGACTGTAGTACTCTTAAGGACACACTGTGTG 918
Qy 38 GlyLysSerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGluLeuMetGly 57
Db 919 CAAGAGGAGGAGAGAGCGCCCAACTC-----CAAGAGGGTGCAGAACGCTCTCAAC 969
Qy 58 ValGlnPheValAspAlaLeuLeuLysGlyGlnMetGluMetAlaLysGlyAlaPhe 77
Db 970 AAAGAGACGGCTGACTCTTG-----GGCAATTGGAGGCTAGCAAGGGAAGCACC 1020
Qy 78 LysThrGlnLeuGluValLeuGluLysValHisProAspGlnPhe----- 92
Db 1021 CGCGAAGTTGAGNAAGATGACCCCATGAATGAGCAGAGAGTTGCCCTCGAAGGAAG 1080
Qy 93 -----AspLysTyzLysLysLeuLysValAspAspLeuAlaAlaAspAlaValMetGln 110
Db 1081 CTTGCCGATGCCAGCAAGAGCTCAAGTTGAAGAAGCTCGTCCGTTGAAATCAACAAG 1140
Qy 111 Gln-----AlaGluMetAlaLysLeuGlnProLysSerGlyAsnAlaPhe 125
Db 1141 CAAAGAAGCTTGTGCGAGCGCGAATGCGCTGACTTGAAGAAGAACTGCCAAGATGTTGAC 1200
Qy 126 IleAspMet-----LeuAsnGlyAsnGlyIleProLysSerSerIleArgGlyLeu 143
Db 1201 CTCTCATTTGAGNAAGTTGAGCCGAGAGAGATGCGAAGGAGCACCAGATTCGTCTCTT 1260
Qy 144 GluAspAlaIleArgThrGlnArgAsp-----MetGluAsnThrAsp 157
Db 1261 CAAGATGAATGAGACAAACAGCAGCAGCAACATCTCTAAGCTTAAACAGGAGCGCAAGAAC 1320
Qy 158 ProSerGluGlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeuProGly 177
Db 1321 CAAGAGAGCAGAACCAAGAACTACCGAGGAT---CTCCAAGCTGTGTGAGGAACAAAC 1377
Qy 178 LeuValAlaAsnMetIleAlaGlyLys----- 186
Db 1378 TTGGCAGCCACACAGCTCAAGGCTAAATCATGCAATCTTGAAGATCTCGAGCAGACC 1437
Qy 187 -----AsnProPheLysMetProGlnGlnMetArgLysAlaGlnAla 200
Db 1438 ATGGAGCGGAGAGAGAAACCGTCTGATATGGACAAGATAAGAGAAAGGCCGAA--- 1494
Qy 201 AlaProSerValPheGlnGlnAlaLeu-----Ala 211
Db 1495 GGAGAGCTCAAGATTGCTCAAGAAACCTTGAAGAGCTCAACAAAGTCAAGAGCGATGCC 1554
Qy 212 GlnArgAlaMetLeuGlyLysAsnAlaProVal-----AlaGlyGlyArgGlyGluGlu 229
Db 1555 GAAATGCTTTGAGAGCGCAGAAACCAATTCACCTTGGGATGAAATCAACTCGAAGAC 1614
Qy 230 GlnArgMetMetMetAsnArgValAspGlnArgMetGlnGlnArgGluLeuGlnGlu 249
Db 1615 GAGCAAGCTGTGTTGCTAACTCAAAAGGAATTCAGCAAGAGCAGGCTAGAGTCAAG 1674
Qy 250 AspGluAspAspAspLeuGluAspGluAspValProArgArgArgSerSerAspGly 269
Db 1675 GATCTT---CATGATCAACTTGGCGATGAGAAGGATGCTCGCAACGTTGCTGATCGCTCG 1731

```

RESULT 13

US-10-956-157-859
; Sequence 859, Application US/10956157
; Publication No. US20050118625A1

```

Qy 270 GluProGlnSerGluAlaGluHisGlnArgArgAspLeuAlaArgArgLeuLysSerSer 289
Db 1732 AGAGCCGACCAACAGGCTGAGTACGAT-----GAGCTCACCAGCACTCGAAGACCAA 1785
Qy 290 ProArgLeuLysGluLeuLeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrGlnArg 309
Db 1786 GCCCGT-----GCCACCGTGCCCAAAATTTGAGTCGCAAAA 1821
Qy 310 MetArgAspSerProLeuSerLysArgArgProLeuAlaMetAsnAspGluAspGluSer 329
Db 1822 AAGAAGGATGCTGAGCTTACCAAGCTCCGT-----CGTGATCTTTGAAGAGTCT 1869
Qy 330 AlaPheArgAlaMetGluAlaArgAlaLysLeuAspGlnLysSerGlnLeuValLeuGly 349
Db 1870 GGACTTAATTCGGAGAGCAACTCAGTCTTCTCAAGAGAG-----GGA 1914
Qy 350 LeuHisGlyPheGlyGluSerAspAspGluAspGluGluAspGluAsnLeuLeuAsp 369
Db 1915 TCAGATGCTATTCAAGAGCTTTCCGATCAAAATCGAGCAA----- 1953
Qy 370 ProSerGluAsnSerPheArgArgAlaProLeuArgLeuSerSerGlyPheValGluLys 389
Db 1954 -----CTTCAAAAGCAAAAGGGAGAAATTGAGAAG 1983
Qy 390 LeuLysSerAsn-----AspGluLeuLysSerAlaLeuAspArg---11e 403
Db 1984 GAGAAAGACATATGCAACGGAGTTCATGAGTCTCTGCTGCTCTTGACCAAGAGCC 2043
Qy 404 LysTyrArgValAspAsp-----ValGluLysTyrLeuAlaProLysProMetGlu-Ph 421
Db 2044 AAGTCCGCGTGCACAGGAGAAATGCAAGGATACGAGGTTCAGACTTCTTTCAGCTT 2103
Qy 421 e-----AsnPro-----LysProGlnProGlyTy 429
Db 2104 CGTCTCAAGGCTGACAACTCCCGCAGCTTCAAGATTTCTGCTCTCTCAAGGGAGCGC 2163
Qy 429 rPheAlaProArgLysIleProThr-----ArgProArgLys 441
Db 2164 CTCAACTCCGAGA---ACTCCGATTTGCCCGCCNAGTCGAAGAAATTGGAGGCCNAGATC 2220
Qy 441 sMetLeuProLeuLeuLeuGlySerAsp---ProLysValGlnGluGluIleArgArgHi 460
Db 2221 CAAGCTTCCAAACCGCTCAAACTCCAAATTTCTCAACGAA--TTGGATCATGCCAAGAGACA 2279
Qy 460 sProSerThrGluTrpLysIleAlaLysGluSerArgValLeuThrAsnLeuLysAsnAs 480
Db 2280 GGCCGAGAGGAA-----TCCCGCAGAGAGCAAAATCTTTTCCAATTTGTGGAAGAA 2330
Qy 480 nProSerLeuAlaAlaLeuPheMetAspLysLeuGluAsnThrLysGlyArgG1 500
Db 2331 CTTGCTCGTGAGCTTGAACAGCTCAAGGAATCATCGAAGACGACGCTCGTGGAAAGAA 2390
Qy 500 nMetLeuThrAspGlu-----GlnLysGlyArgThrArgVa 512
Db 2391 CGAGGCTTCTCGTCAACTTTTCCAAGGCTTCGTTGAACTCGATCAATGGAGAACTAAGTT 2450
Qy 512 LysThrIleArgAlaLeuProArgLeuPheGlyAlaProThrAlaLysAlaGluMetI1 532
Db 2451 CGAGACC-----GAGGGACTCAT 2468
Qy 532 eAspAlaLysValPheGlnAspIleGluGluArgProIleProLeuPhePheGluPr 552
Db 2469 TGGAGCGCAGGAGTTCGATGAGGTCAA--GAACCCCAAAACCAAGAGACTTCGGAGATCC 2527
Qy 552 oLys 553
Db 2528 AAGA 2531

```

```
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
; TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
; FILE REFERENCE: 031896-043000 (AM 101081)
; CURRENT APPLICATION NUMBER: US/10/956,157
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 319805
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 859
; LENGTH: 5883
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-956-157-859

Alignment Scores:
Pred. No.: 4,67e-05 Length: 5883
Score: 162.50 Matches: 136
Percent Similarity: 35.90% Conservative: 88
Best Local Similarity: 21.79% Mismatches: 213
Query Match: 4.94% Indels: 192
DB: 25 Gaps: 25

US-10-736-868-2' (1-643) x US-10-956-157-859 (1-5883)

Qy 12 GlyPheCysIleAlaProLeuSer-----AlaGlnSerProSerThrSerAsp----- 27
Db 3860 GGCTTCTCAGCCAGTCGACGAGCAAGTCCAGCAAGCTCACCAGGACTTCTCCGCGGTGG 3919
Qy 28 AlaProGly-----AlaLeuLeuSerSerLeuValGlyLysSerHisGlnLysLeuPro 45
Db 3920 AGTCCCCAGTCGACGACACTCAGGAGCTG-CTGCAGGAGGAGAACCGGCAGAGCTGAGC 3978
Qy 46 LeuAlaProSerMetGluAlaLeuGlu----- 54
Db 3979 CTGAGCACCAGCTCAGCAGGAGTGAGGAGCAGAGAGNATTCCTTCCGGGAGCAGCTGGAG 4038
Qy 55 -----LeuMetGlyValGlnPhe 60
Db 4039 GAGGAGGAGGAGGAGCAGCAACCTGGAGAAGCAGATCGCCACCCTCCATGCCAGGTG 4098
Qy 61 ValAspAlaLeuLeuLysLysGlyGlnMetGluMetAlaLysGlyAlaPheLysThrGln 80
Db 4099 GCCGAC-----ATGAAAAAGAGATGGAGACAGTGTGGGTGCTTGGAAACTGCT 4149
Qy 81 LeuGluValLeuGluLysValHisProAsp----- 90
Db 4150 GAGGAGGTGAAGGAGAGCTCCAGAGGACCTGGAGGGCCTGAGCCAGCGCAGGAGGAG 4209
Qy 91 -----GlnPheAspLysTyrLysLysLeuLys-----ValAsp 101
Db 4210 AAGGTGGCCGCTACGACAGCTGGAGAGACCAAGACCGGGCTGCACGAGGAGCTGGAC 4269
Qy 102 AspleuAlaAlaAspAlaValMetGlnGlnAlaGluMetAlaLysLeuGlnProLysSer 121
Db 4270 GACCTGTGTGGACCTGGACCAACCGCCAGAGCGCGTGTCAACCTGGAGAGAGACAG 4329
Qy 122 GlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIleProIleGly----- 137
Db 4330 AAGNAG-----TTTGACCACTCTCGCGGAGGAGAGACCATCTCTGCCAAGTATGCA 4383
Qy 138 -----SerSer 139
Db 4384 GAGGAGCGCAGCCGGGCTGAGCGGAGGCCCGAGAGAAGGAGACCAAGGCTCTGTCTGCTG 4443
Qy 140 IleArgGlyLeuGluAspAlaIleArgThrClnArgAspMetGluAsnThrAspProSer 159
Db 4444 GCCCGGCCCTTGGAGGAAGCCATCGGAGCAGAGCGGAGCTGTGGAGCGGCTCAAC----- 4497
Qy 160 GluGlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeuProGlyLeuVal 179
Db 4498 -----AAGCAGTTCCGCACGAG-----ATG 4518

180 AlaAsnMetIleAlaGlyLysAsnProPhe-----LysMetProGlnGlnMetArgLysAla 198
4519 GAGGACCTTATGAGCTTCAAGATGATGTGGCAAGAGATGTCCACGAGTGGAGATGC- 4577
199 GlnAlaAlaProSerSerValPheGlnGlnAlaLeuAlaGlnArgAlaMetLeuGlyLys 218
4578 CAAAGCGGCGCT-----AGACGACGAGTGGAGGA 4607
219 Asn-----AlaProValAlaGlyGlyArgGlyGluGluGlnArgMetMetAsn 235
4608 GATGAAGACGACGCTGGAAGAGCTGGAGACGAGCTGCAGGCCACCAAGATGCCAAGCT 4667
236 -ArgValAspGlnArgMetGln-----GlnArgGluLeuGlnGlu 249
4668 GCGGTTGGAGGTCAACCTCGACGCCATGAAGGCCAGTTCGAGCGGAGCTTCGACGGCCG 4727
249 uAspGluAspAspAspLeuGluAspValProArgArgSerSerAspG1 269
4728 GGACGACGACGAGGAGGAGAAAGACGCTGGTTCAGACAGTCCGGGAGATGGAGGC 4787
269 yGluProGlnSerGluAlaGluHisGlnArgArgAspLeuAlaArgLeuLysSerSe 289
4788 AGAGCTGGAGGACGAGAGGAGCAGCGCTCGATGGCAGTGGCGGCCGGAAGAAGTGA 4847
289 rProArgLeuLysGluLeu----- 295
4848 GATGGACCTGAAAGGACCTGGAGCGGCACATCGACTCGGCCAACAGAACCGGGACGAAGC 4907
296 -----LeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrGlnArgMetArgAs 312
4908 CATCAACAGCTCCGGAAGCTGCGGCCACAGATGAAGGACTGATCGCGGAGTGGATGA 4967
312 pSerProLeuSerLysArgArgProLeuAlaMetAsnAspGluAspGluSerAlaPheAr 332
4968 CACCCGCGCTCTCGTGAGGAGATCTGGCCAGGCCAAAGAGACGAGAGAAGAGTGA 5027
332 gAlaMetGluAlaArgAlaLysLeuAspGlnLysSerGlnLeuValLeuGlyLeuHisG1 352
5028 GAGCATGGAGGCC-----GAGATGATCCAGTTGCAGGAGGAACATGGCAGC 5072
352 yPheGlyGluSerAspAspGluAspGluAspGluAsnLeuIleAspProSerG1 372
5073 CGCGGAGCTGCCAAGCGCCAGCAGGAGCGGATGAGCTGGCTGACGAGATCGC 5132
372 uAsnSerPheArgArgAlaProLeuArgLeuSer----- 383
5133 CAACAGCAGCGCAAGAGAGCCCTGGCGTTAGAGGAGAGCGCGCTCTGGAGGCCCGCAT 5192
384 -SerGlyPheValGluLysLeuLysSer-----AsnAspG1 395
5193 CGCCGAGCTGGAGGAGGAGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5252
395 uLeuLysSerAlaLeuAspArgIleLysTyrArgValAspAspValIclulysTyrLeuAl 415
5253 GCTGAGAAGGCC-----AACCTGCAGATCGACCATC----- 5286
415 aProLysProMetGluPheAsnProLysProGlnProGlyTyrPheAlaProArgLysI1 435
5287 -----AACAC-CGACCTGAACCTGGA----- 5306
435 eProThrArgProArgLysMetLeuProLeuLeuIleGlySerAspProLysValGlnG1 455
5307 -----GGCGACCGACCGCC-AGAAGACGAGAGATGTCGGCAGCAGCTGGAGCAACCGCAAC 5361
455 uGluIleArgArgHisProSerThrGluTyrLysIleAlaLysGluSerArgValLeuTh 475
5362 AAGGAGCTTAAGTCAAGCTGACGAGGATGG-----AGGCGACTGTGC 5403
475 rAsnLeuLysAsnAsnProSerLeuAlaLeu-----PheMetAspAs 490
5404 AAGTCC-AAGTACAAGCGCTTCAATCCCGCTCGAGGCCAAGATTGCACAGCTGGAGGA 5462
```

QY 490 p1ysLeuGluAanThrLeuLysGlyArgGlnMetLeuThrAspGluGlnLysGlyArgTh 510
Db 5463 GCAGCTGGACCAAGAGGAGCGCCAGGAGCGCTGC---AAACAGGTGCGTGGAC 5519
QY 510 rArgValLys 513
Db 5520 CGAGAAGAAG 5529

RESULT 14

US-10-283-975A-279
; Sequence 279, Application US/10283975A
; Publication No. US20040110792A1
; GENERAL INFORMATION:
; APPLICANT: Ortho-Clinical Diagnostics, Inc.
; TITLE OF INVENTION: Methods For Assessing and Treating Leukemia
; FILE REFERENCE: CDS 293 PCT
; CURRENT APPLICATION NUMBER: US/10/283,975A
; CURRENT FILING DATE: 2002-10-30
; PRIOR APPLICATION NUMBER: 60/340,938
; PRIOR FILING DATE: 2001-10-30
; PRIOR APPLICATION NUMBER: 60/338,997
; PRIOR FILING DATE: 2001-10-30
; PRIOR APPLICATION NUMBER: 60/340,081
; PRIOR FILING DATE: 2001-10-30
; PRIOR APPLICATION NUMBER: 60/341,012
; PRIOR FILING DATE: 2001-10-30
; NUMBER OF SEQ ID NOS: 900
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 279
; LENGTH: 3320
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(3320)
; OTHER INFORMATION: N=any base
US-10-283-975A-279

Alignment Scores:
Pred. No.: 2,42e-05 Length: 3320
Score: 162.00 Matches: 160
Percent Similarity: 38.28% Conservative: 135
Best Local Similarity: 20.75% Mismatches: 251
Query Match: 4.92% Indels: 228
Gaps: 39

US-10-736-868-2 (1-643) x US-10-283-975A-279 (1-3320)

QY 15 IleAlaProLeuSerAlaGlnSerProSerThrSerAspAlaProGlyAlaLeuLeuSer 34
Db 656 GTGGCGTCCCTCAGTTCCAGCTCCAGGACACCCAGGAGCTG----- 697
QY 35 SerLeuValGlyLysHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGlu 54
Db 698 ---CTTCAAGAAGAAACCGGAGAGCTCAACGTGTCTACGAAGCTCGCCAGCTGGAG 754
QY 54 ----- 54
Db 755 GAGGAGCGGAACAGCTGCAAGACAGCTGAGGAGGAGAGTGGAGGCCAAGCAGACCTG 814
QY 55 -----LeuMetGlyValGlnPheValAspAlaLeuIleLysLysGlyGln 69
Db 815 GAGCGCCACATCTCCACTCTCAACATCCAGCTCTCCGACTCGAAG---AAGAAGTGCAG 871
QY 70 -----MetGluMetAlaLysGlyAlaPheLysThrGlnLeu 81
Db 872 GACTTTGCCAGCACCGTGAAGCTCTGGAAGAGGGAAGAGAGTTCCAGAAGGAGATC 931
QY 82 GluValLeuGluLysValHisProAspGln-----PheAspLysTyrLysLysLeu 98
Db 932 GAGAACCTCACCCAGCTACGAGGAGAGGCGCGCGGTATGATAAACTGGAAAGACC 991
QY 99 Lys-----ValAspAspLeuAlaAlaAspAlaValMetGlnGlnAla 112

Db 992 AAGAACAGGCTTCAGCAGGAGCTGGACACCTGGTTGTTGTTGTTGGACAACCCAGCGCAA 1051
QY 113 GluMetAlaLysLeuGlnProLysSerGlyAsnAlaPheIleAspMetLeuAsnGlyAsn 132
Db 1052 CTGCTGTCCAACTCGAAAGAGAGCAGAGAAA-----TTTGATCAGTTGTTAGCCGAG 1105
QY 133 GlyIleProIleGlySerSer----- 139
Db 1106 GAGAAAAACATCTCTTCCAAATACCGGATGAGAGGAGCAGAGCTGAGGCAGAGCCAGG 1165
QY 140 -----IleArgGlyLeuGluAspAlaIleArgThrGln 150
Db 1166 GAGAAGGAAACCAAGGCCCTGTCCCTGGCGCCCTTGAAGAGCCCTTGAAGCCAAA 1225
QY 151 ArgAspMetGluAsnThrAspProSerGluGlnIleAlaLysAlaValMetAspLysPhe 170
Db 1226 GAGGAACCTCGAGCGGACCAAC-----AAAATGCTCAAGAGCCGAATGGAA----- 1270
QY 171 GlnThrGlnIleLeuProGlyLeuValAlaAsnMetIleAlaGlyLysAsnProPhe--- 189
Db 1271 -----GACCTGGTCAGCTCCAAGATGACCTGGGC 1300
QY 190 LysMetProGlnGlnMetArgLysAlaGlnAlaProSerSerValPheGlnGlnAla 209
Db 1301 AAGAACGTCCATGAGCTGGAGAGTCCAAAGCGGCGCTGGAGACCAGATGGAGGAGATG 1360
QY 210 LeuAlaGlnArgAlaMetLeuGlyLysAsnAlaProValAlaGlyGlyArgGlyGluGlu 229
Db 1361 AAGCGCAGCTGGAGAGCTGGAGGAGCTGCAAGCCACG-----GAGGAC 1408
QY 230 GlnArgMetMetMetAsnArgValAspGlnArgMetGln-----Gln 243
Db 1409 GCCAAACTG-----CGCTGGAGTCAACATGAGGCGCTCAAGGGCCAGCTTCGAA 1459
QY 244 ArgGluLeuGlnGluAspGluAspAspAspLeuGluAspGluAspValProArg 263
Db 1460 AGGATCTCCAAAGCCCGGACGAGCAATCAG-----GAGAAGAGGAGCAACTGCAG 1513
QY 264 ArgArgSerSerAspGlyGluProGlnSerGluAlaGluHisGlnArgArgAspLeuAla 283
Db 1514 AGACAGCTTCACGAGTATGACACGGAACCTGGAAGACGAGCGAAGCAACGTGCCCTGGCA 1573
QY 284 ArgArgLeuLysSerSerProArgLeuLysGluLeuLeuGlnAsnAlaGluValGlnSer 303
Db 1574 GCTCAGCAAAAGAAG-----AAGCTGGAAGGAGACCTGAAAGACCTCGAGCTTCAGGCC 1627
QY 304 LeuLeuSerTyrGlnArgMetArgAspSerProLeuSerTysArgArgProLeu----- 321
Db 1628 ---GACTCTGCCATCAAGGGGAGGAGGAGCCATCAAGCAGCTACGCAAACTGCAGGCT 1684
QY 322 AlaMetAsnAsp-----GluAspGluSerAlaPheArg---AlaMetGlu 335
Db 1685 CAGATGAAGGACTTTCAAAGAGAGCTGGAAGATCCCGCTGCCAGAGATGATGATCTTT 1744
QY 336 AlaArgAlaLysLeuAspGlnLysSerGlnLeuValLeuGlyLeuHisGlyPheGlyGlu 355
Db 1745 GCCACAGCAAGAGATGAGAAAGAACCCAGAGCTTG-----GAA 1786
QY 356 SerAspAspAspGluAspGluAspGluAsnLeuIleAspProSerGluAsnSerPhe 375
Db 1787 GCAGACCTCATGACGTACAAAGAGGAGC-----CTCGCGCGCGCTGAGAGGGCTCGC 1837
QY 376 ArgArgAlaProLeu-----ArgLeuSerSerGlyPheValGluLysLeuLysSer 392
Db 1838 AAAACAGCGGACCTTCGAAAGGAGGAGAACTGGCAGAGGAGCTGGCCAGTAGCTGTGCGGA 1897
QY 393 AsnAspGluLeuLysSerAlaLeuAspArgIleLysTyrArgValAspValGluLys 412
Db 1898 AGGAACGCACTCCAGGAGGAGAGCGCGCTGGAGGCCCGGATCGGCCAGCTGGAGAG 1957
QY 413 TyrLeuAlaProLysProMetGlnPheAsnProLysProGlnProGlyTyrPheAlaPro 432

Db 1958 GAGCTGAGGAGGAGCAGGG-CAACATGGAGCCATGAGCGACCGGGTCC-GCAAGGCCA 2015
Qy 433 ArgLysIleProThr-----ArgProArgLys 441
Db 2016 CACAGCAGGCGGAGCAGCTCAGCAACGAGCTGGCCACAGAGCGCAGCGCCGAGGA 2075
Qy 442 Met---LeuProLeuLeuIleGlySerAspProLysValGlnGluGluIleArgArgHis 460
Db 2076 ATGAGAGTTCCC-----GGCAGCAGCTTGGCGGCAGAACAGAGAGCTCCGAGCA 2126
Qy 461 ProSerThrGluTrpLysIleAlaLysGluSerArgValLeuThrAsnLeuLysAsnAsn 480
Db 2127 AGCTCCAGAGATGG-----AGGGGGCGGTCAAGTCCCAAGTTCAAT 2168
Qy 481 ProSerLeuAlaLeu-----PheMetAspAspLysLeuGluAsnThr 495
Db 2169 CCACC-ATCGCGCGCTGGAGGCCAAGATTGCACAGCTAGAGGAGCAGGTTCGAGCAGGAG 2227
Qy 496 LeuLysGlyArgGlnMetLeuThr----- 503
Db 2228 GCCAGAGAGAACAGCGCGCCACCAAGTCGCTGAGCAGAGAACAGAACAGAGCTGAAGGAA 2287
Qy 504 -----AspGlu-----Gln 506
Db 2288 ATCTTGCTGAGGTGGAGGAGCGCGCAAGATGCGCCGAGCAGTACAGAGGAGCAGCAGAG 2347
Qy 507 LysGlyArgThrArgValLysThrIleArgAlaLeuProArgLeuPheGlyAlaProThr 526
Db 2348 AAAGGCAATCGAGGTCAAGCAGCTCAAG-----AGGCAGCTGGAGGAGCAGAG 2398
Qy 527 AlalysAlaGluMetIleAspAla-----LysValPheGlnAspIleGluGluArg 543
Db 2399 GAGAGTCCCGAGCGCATCAACGCCAACCGCAGGAAGCTGCGCGGAGCTGGATGAGGCC 2458
Qy 544 ProfileProLeuPhePheGluProLysGlyArgHis-----ThrArg 558
Db 2459 ACGAGAGCAAC-----GAGGCCATGGCGCGCGAGGTGAACGCACCTCAAGAGCAAG 2509
Qy 559 LeuArgTrpThrGlyAlaAsnGluLysGluIleProGly-----LeuGlySerArgPhe 576
Db 2510 CTCAGG---CGAGAAACGAGACCTCTTTCGTTCTCTAGAAGGTCTGGAGGACGTAGA 2566
Qy 577 IleLeuProSerLeuAspProThrMetProAlaLeuAsnThr----- 590
Db 2567 GTTATTGAAATGCAGATGTTCTTGAGGAGAAACGAGCACTCGAGCGCAGACTTCAAT 2626
Qy 591 -----AlaPheSerThrGlnGlyArgAlaArg 599
Db 2627 GGAACACAGGCCAGTGAATAAGCAACTTCTACAGTTTTCACCCAGCGCAAGAAACCAA 2686
Qy 600 AspGluTrpAspThrMetPheLysIleProAsnAsnTrpAsnProGlyAspGluValGly 619
Db 2687 AAACCAAAACAAACAAACAAACAAACAAACAAAC-----AACCCA----- 2728
Qy 620 PheLysMetAsnSerLysThrLysArgPheVal 630
Db 2729 -----GAACAAACGAAACCCAGCAGACTGTA 2755

RESULT 15
US-09-954-456-1602
; Sequence 1602, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20

; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1602
; LENGTH: 3388
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-954-456-1602

Alignment Scores:
Pred. No.: 4,94e-05 Length: 3388
Score: 159.00 Matches: 150
Percent Similarity: 38.30% Conservative: 130
Best Local Similarity: 20.52% Mismatches: 230
Query Match: 4.83% Indels: 225
DB: 9 Gaps: 35

US-10-736-868-2 (1-643) x US-09-954-456-1602 (1-3388)

Qy 15 IleAlaProLeuSerAlaGlnSerProSerThrSerAspAlaProGlyAlaLeuLeuSer 34
Db 1270 GTGGCGTTCCTCAGTTCCTCCAGCTCCAGGACACCCAGGAGTTG----- 1311
Qy 35 SerLeuValGlyLysSerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGlu 54
Db 1312 ---CTTCAAGAAAGAAACCCGCGCAGAGCTCAACGTGTCTACGAAGCTGCGCCAGCTGGAG 1368
Qy 54 ----- 54
Db 1369 GAGGAGCGGAACAGCTCGCAAGACCAGCTGGAGCAGAGATGGAGCCACGAGAACCTG 1428
Qy 55 -----LeuMetGlyValGlnPheValAspAlaLeuIleLysLysGlyGln 69
Db 1429 GAGCGCCACATCTCCACTCTCAACATCCAGCTCTCCGACTCGAAG---AAGAAAGCTGCAG 1485
Qy 70 -----MetGluMetAlaLysGlyAlaPheLysThrGlnLeu 81
Db 1486 GACTTTGCCAGCACCGTGGAGAGCTCTGGAAGAGGGGAAAGAGGTTCCAGAGAGGAGATC 1545
Qy 82 GluValLeuGluLysValHisProAspGln-----PheAspLysTyrLysLysLeu 98
Db 1546 GAGAACTTACCAGCAGCTACGAGGAGAGAGCGCGCTTATGATTAACCTGGAAGAGACC 1605
Qy 99 Lys-----ValAspAspLeuAlaAlaAspAlaValMetGlnGlnAla 112
Db 1606 AAGAACAGCGCTTCAGCAGGAGCTGGAGCAGCTGTTGTTGTTGATTGGACAACACGCGCAA 1665
Qy 113 GluMetAlaLysLeuGlnProLysSerGlyAsnAlaPheIleAspMetLeuAsnGlyAsn 132
Db 1666 CTCGTGTCCAACTCGAAAGAAAGACAGAGGAAA-----TTTGATCAGTTGTTAGCCGAG 1719
Qy 133 GlyIleProIleGlySerSer----- 139
Db 1720 GAGAAAAACATCTCTTCCAAATACGCGGATGAGAGGAGCAGAGCTGAGGCGAGACCGAG 1779
Qy 140 -----IleArgGlyLeuGluAspAlaIleArgThrGln 150
Db 1780 GAGAAAGGAAACCAAGGCCCTGCTCCCTGGCTCGGGCCCTTGAAGAGGCTTGAAGAGCCAA 1839

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 18, 2005, 12:25:31 ; Search time 5597 Seconds
(without alignments)
4372.936 Million cell updates/sec

Title: US-10-736-868-2

Perfect score: 3291

Sequence: 1 MILFLFLLLGFCIAPLSA.....SKTRFVGNGAFDMPALGL 643

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/csm2_1/USPTO spo01/US10736868/runat 14072005 105510 2290/app_query.fasta_1.839
-DB=EST -QFMT=fastap -SURFIX=src -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10736868 @CGN 1.1 3437 @runat 14072005 105510 2290 -NCPU=6 -ICPU=3
-NO.MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*

1: gb_est1: *
2: gb_est2: *
3: gb_hic: *
4: gb_est3: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_gse1: *
9: gb_gse2: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	484	14.7	300	1 AV178298	AV178298 AV178298
2	349	10.6	476	4 B1741875	B1741875 kt82b03.Y
3	167	5.1	3699	9 AY421012	AY421012 Pan trogl
4	162.5	4.9	5883	9 AY420735	AY420735 Homo sapi
5	162.5	4.9	5883	9 AY420737	AY420737 Mus muscu
6	159	4.8	4026	9 AY421011	AY421011 Homo sapi
7	155.5	4.7	2993	3 AK030316	AK030316 Mus muscu
8	153.5	4.7	7139	3 BC050866	BC050866 Mus muscu
9	153.5	4.7	7139	3 BC050868	BC050868 Mus muscu

10	152.5	4.6	3918	9 AY421013	AY421013 Mus muscu
11	152	4.6	2934	9 AY414528	AY414528 Homo sapi
12	151	4.6	3012	9 CL960352	CL960352 OaIFCC004
13	150.5	4.6	1128	3 CR679256	CR679256 Tetraodon
14	150.5	4.6	3431	3 AK030449	AK030449 Mus muscu
15	150	4.6	837	5 BW470454	BW470454 BW470454
16	150	4.6	2740	3 AK029960	AK029960 Mus muscu
17	150	4.6	2928	9 AY414530	AY414530 Mus muscu
18	148	4.5	3636	3 AK029997	AK029997 Mus muscu
19	148	4.5	4613	3 AK033263	AK033263 Mus muscu
20	147.5	4.5	647	3 CNS09FH4	BX056132 Single re
21	147.5	4.5	6275	3 AK029236	AK029236 Mus muscu
22	146.5	4.5	3006	9 AY409299	AY409299 Homo sapi
23	146.5	4.5	3803	3 HSM801462	AL133606 Homo sapi
24	146.5	4.5	4277	3 CR627431	CR627431 Homo sapi
25	146	4.4	5504	9 AY407955	AY407955 Homo sapi
26	145.5	4.4	685	5 BW034508	BW034508 BW034508
27	145.5	4.4	715	5 BW444409	BW444409 BW444409
28	144.5	4.4	1168	6 CD506303	CD506303 CDA78-E05
29	144.5	4.4	4064	3 AK033354	AK033354 Mus muscu
30	144.5	4.4	5495	9 AY407957	AY407957 Mus muscu
31	144	4.4	1237	5 BU510057	BU510057 AGENCOURT
32	144	4.4	4269	3 AK038318	AK038318 Mus muscu
33	144	4.4	6294	9 AY410631	AY410631 Homo sapi
34	143.5	4.4	3191	3 CR623413	CR623413 full-length
35	143	4.3	999	4 BM477724	BM477724 AGENCOURT
36	143	4.3	3395	3 BC057135	BC057135 Mus muscu
37	142	4.3	3914	3 BC032177	BC032177 Mus muscu
38	141.5	4.3	880	9 CG809737	CG809737 FSAAF38TF
39	141.5	4.3	3006	9 AY409300	AY409300 Pan trogl
40	141.5	4.3	3981	9 CL964889	CL964889 OaIFCC011
41	141	4.3	708	7 CK659858	CK659858 LP17954.5
42	141	4.3	771	7 CK478311	CK478311 AGENCOURT
43	141	4.3	2358	3 AK050669	AK050669 Mus muscu
44	141	4.3	3286	3 AK030978	AK030978 Mus muscu
45	141	4.3	4400	3 BC053625	BC053625 Homo sapi

ALIGNMENTS

RESULT 1
AV178298/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

AV178298 300 bp mRNA linear EST 21-JUL-1999
AV178298 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
embryo Caenorhabditis elegans cDNA clone yk563c9 3', mRNA sequence.

AV178298

AV178298.1 GI:5558199

EST.

Caenorhabditis elegans

Caenorhabditis elegans

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;

Rhabditoidea; Rhabditidae; Pelodierinae; Caenorhabditis.

1 (bases 1 to 300)

Kohara, Y., Shin-i, T., Thierry-Mieg, J., Thierry-Mieg, D., Mitauki, H.,

Nishigaki, A., Motohashi, T., Zeng, Q., Watanabe, H., Sugimoto, A.,

Sano, M., Miyata, A., Mitani, Y., Iida, K., Uesugi, H., Sugiyama, Y. and

Nomoto, H.

Expressed genes in C.elegans

Unpublished (1999)

Contact: Yuji Kohara

Genome Biology Lab.

National Institute of Genetics

Yata 1111, Mishima, Shizuoka 411, Japan

Tel: 81-559-81-6854

Fax: 81-559-81-6855

Email: ykohara@lab.nig.ac.jp.

Location/Qualifiers

1. 300

/organism="Caenorhabditis elegans"

/mol_type="mRNA"

/strain="N2"

/db_xref="taxon:6239"

/clone="yk563c9"

```

/sex="hermaphrodite"
/dev_stage="embryo"
/clone_lib="Yuji Kohara unpublished cDNA:Strain N2
hermaphrodite embryo"

ORIGIN
Alignment Scores:
Pred. No.: 8.99e-38 Length: 300
Score: 484.00 Matches: 88
Percent Similarity: 97.78% Conservative: 0
Best Local Similarity: 97.78% Mismatches: 2
Query Match: 14.71% Indels: 0
DB: 1 Gaps: 0

US-10-736-868-2 (1-643) x AV178298 (1-300)
Qy 554 GYVARGHISThrArgLeuArgTTPThrGlyAlaAsnGluLysGluIleProGlyLeuGly 573
Db 298 GGAAGGCATACGAGATTGAGATGCTGGAGCAAAATGAAAGAAATTCAGGACTTGA 239

Qy 574 SerArgPheLeuLeuProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSer 593
Db 238 AGTCGCTTCATCTCCCATNTCTTGATCCAACTATGCCANCTTGAACACGGCTTCTCG 179

Qy 594 ThrGlnGlyArgAlaArgAspGluTTPAspThrMetPheLysIleProAsnAsnTTPAsn 613
Db 178 ACTCAGGGCGAGCCGCTGACGAGTGGGATACCATGTGTTCAAAATCCGGAATACTGGAAT 119

Qy 614 ProGlyAspGluValGlyPheLysMetAsnSerLysThrLysArgPheValGlyGlyAsn 633
Db 118 CTGGAGATGAAGTTGGTTCAAAATGAATCAAAACCAACAGATTCGTGGAGGAAT 59

Qy 634 GlyAlaPheAspMetProAlaLeuGlyLeu 643
Db 58 GGAGCATTTGATATGCTCGCACTGGGATTG 29

RESULT 2
BI741875 476 bp mRNA linear EST 25-SEP-2001
LOCUS Strongyloides ratti L1 pAMP1 v3 Chiapelli McCarter
DEFINITION Strongyloides ratti cDNA 5' similar to TR:P91120 P91120 SIMILAR TO
HMG-BOX SINGLE-STRANDED DNA AND RNA BINDING PROTEINS. [1] ;, mRNA
sequence.
ACCESSION BI741875.1 GI:15742831
VERSION Strongyloides ratti
SOURCE Strongyloides ratti
ORGANISM Strongyloides ratti
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Panagrolaimidae; Strongyloidiidae; Strongyloides.
McCarte, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J.,
Wylie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B.,
Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Franklin, C.,
Tsagaris, N., R., Ronko, I., Kennedy, S., Maguire, L., Beck, C.,
Harwood, K., Stepcoe, M., Allen, M., Person, B., Swaller, T.,
Harvey, N., Schurk, R., Kohn, S., Shin, T., Jackson, Y., Cardenas, M.,
McCann, R., Waterston, R., and Wilson, R.
The Washington Univ. Nematode EST Project, 1999
Unpublished (1999)
Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was constructed by Brandi Chiapelli and Dr. James
McCarter (bchiapelli@watson.wustl.edu & jmcarter@watson.wustl.edu) at
Washington University, St. Louis. DNA Sequencing by: Washington
University Genome Sequencing Center St. Louis.
Seq primer: -40RP from Gibco
High quality sequence stop: 306.

FEATURES
source
Location/Qualifiers
1. .476
/organism="Strongyloides ratti"
/mol_type="mRNA"
/db_xref="taxon:34506"
/dev_stage="L1"
/lab_host="DH10B"
/clone_lib="Strongyloides ratti L1 pAMP1 v3 Chiapelli
McCarter"
/note="Vector: pAMP1 (Gibco); The library was constructed
by Brandi Chiapelli and Dr. James McCarter at Washington
University, St. Louis. The cDNA was made by using
Dynabead oligo-dT priming (Dyna). PCR based library
using a modified protocol from the SMART PCR cDNA
Synthesis Kit from Clontech. Directionally cloned into the
UDG sites of pAMP1. Nematodes were provided by Dr. Mark
Viney of Bristol, UK."

ORIGIN
Alignment Scores:
Pred. No.: 4.02e-24 Length: 476
Score: 349.00 Matches: 76
Percent Similarity: 60.12% Conservative: 22
Best Local Similarity: 46.63% Mismatches: 51
Query Match: 10.60% Indels: 14
DB: 4 Gaps: 4

US-10-736-868-2 (1-643) x BI741875 (1-476)
Qy 483 LeuAlaAlaLeuPheMetAspLysLeuGluAsnThrLeuLysGlyArgGlnMetLeu 502
Db 3 TTAGTTGCTTTATTTGAAGAACTGATCTGGCTGAAAAAATTGATTAACCATCATTTATG 62

Qy 503 ThrAspGluGlnLysGly-----ArgThrArgValLysThrIleArgAlaLeuProArg 520
Db 63 TCAATGAACAAAGAGGTTTGTAGTGTCACAGTGGAAATGCAATTACAACATATCCAGT 122

Qy 521 LeuPheGlyAlaProThrAlaLysAlaGlu-----MetIleAspAlaLysValPheGln 538
Db 123 ATGTATGCTGCTAAATATTATTGTTGATGAACCAAAATATTATGATTAATAACAAACAA 182

Qy 539 AspIleGluArgProIleProLeuPhePheGluProLysGlyArgGlnThrArg 558
Db 183 ATTGTTGAGAAAGAGAAATTCCTCCTTTTGTACCAAAAGAAACATACAGA 242

Qy 559 LeuArgTTPThrGlyAlaAsnGluLysGluIleProGlyLeuGlySerArgPheLeu 578
Db 243 TTAAGGTGGTAAACAGCAACAGAAATTCGCGAATTCGAAAGTAGATTATTATATA 302

Qy 579 ProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSerThrGlnGlyArgAla 598
Db 303 CCATCTCTGATCCAACTAGCCAGCAATTAATTCAGTTGTATCAACACAGGAAAGAA 362

Qy 599 ArgAspGluTTPAspThrMetPheLysIleProAsnAsnTTPAsnProGly----- 615
Db 363 AGAAATGAATATGAACAACCTTGGAAATACCAAAATGCTTGGAAATCTGGAATATTTTG 422

Qy 616 -----AspGluValGlyPheLysMetAsnSerLysThrLysArgPheValGlyGlyAsn 633
Db 423 GAATTCGATGATACAAATCTGAAATTAAT-----TGTTGGGAGCGGAAT 467

Qy 634 GlyAlaPhe 636
Db 468 AGTAAATTT 476

RESULT 3
AY421012 3699 bp DNA linear GSS 17-DEC-2003
LOCUS Pan troglodytes ACINUS gene, VIRUTAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY421012
VERSION AY421012.1 GI:39776969
KEYWORDS GSS.

```

```

SOURCE ORGANISM
Pan troglodytes (chimpanzee)
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE
1 (bases 1 to 3699)
AUTHORS
Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarival, A.,
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
Adams, M.D. and Cargill, M.
TITLE
Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL
Science 302 (5652), 1960-1963 (2003)
PUBMED
14671302
REFERENCE
2 (bases 1 to 3699)
AUTHORS
Clark, A.G., Glanowski, S., Nielson, R., Thomas, P., Kejarival, A.,
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Ferreira, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
Adams, M.D. and Cargill, M.
TITLE
Direct Submission
JOURNAL
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT
This sequence as made by sequencing genomic exons and ordering them
based on alignment.
FEATURES
source
Location/Qualifiers
1..3699
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>3699
/gene="ACINUS"
/locus_tag="HCM7415"
gene
ORIGIN
Alignment Scores:
Pred. No.: 8,34e-05 Length: 3699
Score: 167.00 Matches: 115
Percent Similarity: 36.79% Conservative: 66
Best Local Similarity: 23.37% Mismatches: 176
Query Match: 5.07% Indels: 136
DB: 9 Gaps: 23
US-10-736-868-2 (1-643) x AY421012 (1-3699)
Qy 23 ProSerThrSerAspAlaProGlyAlaLeuLeuSer---SerLeuValGlyLysSerHis 41
Db 157 CCGAGTACCTCAGCAAGATCGCGAGCTGGAGGAGTGACTCTGGACGGGAGCCTCTT 216
Qy 42 GlnLysLeuProLeuAlaProSerMetGluAlaLeuGluLeuMetGlyValGlnPheVal 61
Db 217 CAGGCGTGTGGGTGACCGACCTGAAGCCGCGACTGGAGCAGCGAGGC--- 264
Qy 62 AspAlaLeuIleLysLysGlyGln-----MetGluMetAlaLysGly----- 75
Db 265 -----CTAGCCAGAGCGGGCAGAGAGTGCCTGTCAAGCGGCTCAAGGGGGCTCTA 318
Qy 76 -----ATGCTAGAAAATTACAGAAACACTCAACACCCCATCTGCTTCCAGCCAAATCCAG 378
Qy 83 ValLeuGluLysValHisProAspGlnPheAspLysTyrLysLysLeuLysValAspAsp 102
Db 379 ATTGGTGGAGAAATGAGCAGACAGTTCATAAAACAGATATCTGGAAGAACAGCAGGAG 438
Qy 103 LeuAlaAlaAspAlaValMetGlnAla---GluMetAlaLysLeuGlnProLysSer 121
Db 439 CTACTTAGCGAGCGTCTGGAACGTCGAAGCTCGAAGAGCTGCAAGAACTGAAGAGCTTCA 498
Qy 122 GlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIle-----Pro 135
Db 499 GCTGAGTCGGAGGACGAGATGATCCATCCTCAGGAGTGCGCTTCCCTGCTCCTTGAC 558
Qy 136 IleGlySerSerIleArgGlyLeuGluAspAlaIleArgThrGlnArgAspMetGluAsn 155
Db 1363 -----CCTAATACAGAT 1374

```

RESULT 4
AY420735
LOCUS

AY420735 5883 bp DNA linear GSS 17-DEC-2003

```

DEFINITION Homo sapiens MYH9 gene, VIRTUAL TRANSCRIPT, partial sequence,
GENOMIC survey sequence.
ACCESSION AY420735
VERSION AY420735.1 GI:39776692
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 5883)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 5883)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence as made by sequencing genomic exons and ordering them
based on alignment.
FEATURES
source location/Qualifiers
1..5883
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>5883
/gene="MYH9"
/locus_tag="HCM7323"

gene
1..5883
/locus_tag="HCM7323"

ORIGIN
Alignment Scores:
Pred. No.: 0.00046 Length: 5883
Score: 162.50 Matches: 136
Percent Similarity: 35.90% Conservative: 88
Best Local Similarity: 21.79% Mismatches: 213
Query Match: 4.94% Indels: 192
DB: 9 Gaps: 25

US-10-736-868-2 (1-643) x AY420735 (1-5883)
Qy 12 GlyPheCysIleAlaProLeuSer-----AlaGlnSerProSerThrSerAsp----- 27
Db 3860 GGCTTCTCAGCCAGTCCGACAGCAGAGTCCAGCAAGCTCACCAGGACTTCTCCGCGCTGG 3919
Qy 28 AlaProGly-----AlaLeuLeuSerSerLeuValGlyLysSerHisGlnLysLeuPro 45
Db 3920 AGTCCCAGCTGCAGGACACTCAGGAGGTG-CTGACGAGGAGGAGAACCGCAGAGCTGAGC 3978
Qy 46 LeuAlaProSerMetGluAlaLeuGlu----- 54
Db 3979 CTGAGCACCAGCTCAAGCAGGTGGAGGAGCAGAGAAATTCCTTCGGGAGCAGCTGGAG 4038
Qy 55 -----LeuMetGlyValGlnPhe 60
Db 4039 GAGGAGGAGGAGGAGCAGCAGCACTGGAGAGCAGATCGCCACCTCCATGCCCCAGGTG 4098
Qy 61 ValAspAlaLeuIleLysLysGlyGlnMetGluMetAlaLysGlyAlaPheLysThrGln 80
Db 4099 GCCGAC-----ATGAAAAGAAGATGGAGACAGTGTGGGTGCTTGGAAACTGCT 4149
Qy 81 LeuGluValLeuGluLysValHisProAsp----- 90
Db 4150 GAGGAGGTGAAGAGGAAGCTCCAGAGGACCTGGAGGGGCTTGAGGGCCAGCGGACGAGGAG 4209
Qy 91 -----GlnPheAspLysTyrLysLysLeuLys-----ValAsp 101

```

```

Db 4210 AAGTGGCGCCCTTACGACAGCTGGGAGAGACCAAGACCGCGGTGCGACGAGGAGCTGGAC 4269
Qy 102 AspLeuAlaAlaAspAlaValMetGlnGlnAlaGluMetAlaLysLeuGlnProLysSer 121
Db 4270 GACCTGCTGGTGGACCTGGACCAACAGCGCGGTGCAACCTGGAGAGAAGACGAG 4329
Qy 122 GlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIleProIleGly----- 137
Db 4330 AAGAAG-----TTTGACCAGCTCTGCGCGGAGGAGAACCATCTCTGCCAAGTATGCA 4383
Qy 138 -----SerSer 139
Db 4384 GAGGACGCGACCGGCTGAGCGGAGGCCGAGAGAGGAGACCAAGGCTCTGTGCTG 4443
Qy 140 IleArgGlyLeuGluAspAlaIleArgThrGlnArgAspMetGluAsnThrAspProSer 159
Db 4444 GCCCGGGCCCTGGAGGAGCATGTGGAGCAGAGACGCGGAGCTGGACGGCTCAAC----- 4497
Qy 160 GluGlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeuProGlyLeuVal 179
Db 4498 -----AAGCAGTTCCGACCGAG-----ATG 4518
Qy 180 AlaAsnMetIleAlaGlyLysAsnProPhe---LysMetProGlnGlnMetArgLysAla 198
Db 4519 GAGGACCTTATGAGTCCAGAGGATGATGTGGGCGACAGGTGTCACGAGCTGGAGAGTC- 4577
Qy 199 GlnAlaAlaProSerSerValPheGlnGlnAlaLeuAlaGlnArgAlaMetLeuGlyLys 218
Db 4578 CRAAGCGGCCCT-----AGAGCAGCAGGTGGAGGA 4607
Qy 219 Asn-----AlaProValAlaGlyGlyArgGlyGluGluGlnArgMetMetMetAsn 235
Db 4608 GATGAAGACGCGAGCTGGAAGAGCTGGAGGACGAGCTGCAGCGCCACCGAAGATGCCAAGCT 4667
Qy 236 -ArgValAspGlnArgMetGln-----GlnArgLysLeuGlnGlu 249
Db 4668 GCGTTTGGAGGTCACTCGAGGCCATGAAGGCCAGTTCGAGCGGGAGCTTCGAGGGCCG 4727
Qy 249 uAspGluAspAspAspAspLeuGluAspValProArgArgArgSerSerAspG1 269
Db 4728 GGACGAGCAGAGCGAGGAGAGAGAGCAGCTGGTTCAGAGCTGCGGAGATGGAGGC 4787
Qy 269 yGluProGlnSerGluAlaGluHisGlnArgArgAspLeuAlaArgLeuLysSerSe 289
Db 4788 AGAGCTGGAGGACGAGAGGAGCAGCGCTCGATGGCAGTGGCGCCCGGAAGAAGTGGGA 4847
Qy 289 rProArgLeuLysGluLeu----- 295
Db 4848 GATGGACCTGAAGGACTTGGAGGCTCAGCTCGACTCGGCCCAACAGAACCGGGACGAAGC 4907
Qy 296 -----LeuGlnAsnAlaGluValGlnSerLeuSerTyrGlnArgMetArgAs 312
Db 4908 CATCAAAACAGCTGCGGAAGCTGCAGGCCCATGATGAAGGACTGTCATGCGGAGTGA 4967
Qy 312 pSerProLeuSerLysArgArgProLeuAlaMetAsnAspGluAspGluSerAlaPheAr 332
Db 4968 CACCCGCGCTCTCTGAGAGAGATCTGCGCCAGGCCAAAGAGACGAGAAAGAGCTGAA 5027
Qy 332 gAlaMetGluAlaArgAlaLysLeuAspGlnLysSerGlnLeuValLeuGlyLeuHisG1 352
Db 5028 GAGCATGGAGGCC-----GAGATGATCCAGTTGCAGGAGGAACCTGGCAGC 5072
Qy 352 yPheGlyGluSerAspAspAspGluAspGluGluAspGluAsnLeuIleAspProSerG1 372
Db 5073 CGCGGAGCTGCCAAGCGCCAGCGGAGCGGATGAGCTGGCTGACGAGATCGC 5132
Qy 372 uAsnSerPheArgArgAlaProLeuArgLeuSer----- 383
Db 5133 CAACACGACGCGCAAGAGGAGCCCTGCGTTAGAGGAGAAAGCGCGCTCTGGAGGCCCGCAT 5192
Qy 384 -SerGlyPheValGlyLysLeuLysSer-----AsnAspG1 395

```



```

QY 156 ThrAspProSerGluGlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeu 175
DB   :|||:|||||:
619 TCCTCAATTTCTCAAGAG-----
QY 176 ProGlyLeuValAlaAsnMetIleAlaGlyLysAsnProPheLysMetProGlnGlnMet 195
DB   :|||:|||||:
637 -----AAGGTGACTCTGATGATGAAGAACCA-----
QY 196 ArgLysAlaGlnAlaAProSerValPheGlnGlnAlaLeuAlaGlnAlaMet 215
DB   :|||:|||||:
664 AGCAAGAGAGAGACGATCATCTAGGTC--AGACAGCGAAGAGCAGCTAACTGTCT 720
QY 216 LeuGlyLysAsnAlaProValAlaGlyArgGly-GluGlu-----GlnArgMetMe 233
DB   :|||:|||||:
721 GAGGCG--AGCCAACCTGCTGA-GGAGGAGAGGATCAAGAAACACCTTCCAGAAACCT 776
QY 233 tMetAsnArgValAspGlnArgMet-----GlnGlnArgGluLeuGlnGluAspG 251
DB   :|||:|||||:
777 AAGGTCAGAGCAGATCGAAATTTGAAACAGAGGAGGAAGAGGAGGAGGAGGAGGA 836
QY 251 uAspAspAspLeuGluAspGluAspValProArgArgSerSerSerAspGlyGluPr 271
DB   :|||:|||||:
837 GGAAGAGATGATGAAGAGAGAA-----GGTGATGATGAGG 875
QY 271 oGlnSerGluAlaGluHisGlnArgAspLeuAlaArgArgLeuLysSerSerProAr 291
DB   :|||:|||||:
876 ACAAAATCTAGGAG-----GCACCAAT 899
QY 291 gLeuLysGluLeuLeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrGlnArgMetAr 311
DB   :|||:|||||:
900 CCTGAAGAGTTTAAGGAAGAGGGGAA-----GAGATACCTAGAGTAA 944
QY 311 gAspSerProLeuSerLysArgArgProLeuAlaMetAsnAspGluAspGlu 328
DB   :|||:|||||:
945 ACCAGAGAGATGATGGATGAGAGACCCAAACACAGATCCCGAGAACAGAGGTGTAGA 1004
QY 329 -----SerAlaPheArgAlaMetGluAlaArgAlaLysLeuAspGlnLys 346
DB   :|||:|||||:
1005 GAGAGGAGGAGATTACAGATCCCGAGAGGCTAGA-----AAAAGTCATCT 1055
QY 346 uValLeuGlyLeuHisGlyPheGlyGluSerAspAspGluAspGluAspGluAs 366
DB   :|||:|||||:
1056 GGCC-----AGACAGCAGCAGGAGAGAGGAATGAA 1085
QY 366 nLeuLeuAspProSerGluAsnSerPheArgArgAlaProLeuArgLeuSerSerGlyPh 386
DB   :|||:|||||:
1086 AACACATCTCCCTTGAGGAGGAGAAAGAGAA-----ATAAATCTTCCAAAGGCTT 1139
QY 386 eValGluLysLeuLysSer-----AsnAspGluLeuLysSerAl 399
DB   :|||:|||||:
1140 AAAGGAAAAATCGAAGTCTCTTCCCTCCCTCGACTGACTGAAGATCGAAAGAGGCCCTC 1199
QY 399 aleuAspArglleLysTyrArgValAspValAspValGluLysTyrLeuAlaProLysProMe 419
DB   :|||:|||||:
1200 ACTTGAGCGCTGCCAGAGCAAACTGCCAGCGAGGAGGAG-----ACTCCTCCACCTTT 1253
QY 419 t-----GluPheAsnProLysProGlnProGlyTyrPheAlaProArgLysI 435
DB   :|||:|||||:
1254 ACTAAACAAGAGGAGCATCTTCTCCACCATCTATCCACAGCTCCATCGGAAGAAGAAAT 1313
QY 435 e-----ProThrArgPro----- 439
DB   :|||:|||||:
1314 AGAGCCCATGGAAGGCCCGCCCTGCTCTCATTCAGTTATCTCTCTCTCTCTCTCTCTCT 1373
QY 440 -----ArgLysMetLeu----- 443
DB   :|||:|||||:
1374 TGCTGACACCGAGGAGCTATTAGTATCTCAGCATACTCTCCAGTTGGTAGGAGCCTGTC 1433
QY 444 -ProLeuLeuLeuGlySerAsp-ProLys-----ValGlnGluLeuLeuArgGH 460
DB   :|||:|||||:
1434 TCCITTTGTCAGTCTCTTCCAGACACAAAGAGAGATCTCCAGCAGAAAGTCCAGGAGGA 1493
QY 460 isProSerThrGluTrp-----LysIleAlaLysGluSerArgValLeuT 475

```

```

DB 1494 GAGTGTCTGCTGCTCTGGTTTCAGAAAAGCACACTGGCTGCTACTACTCAGCCCGAGAGGATCT 1553
QY 475 hrAsnLeuLysAsn-----AsnProSerLeuAlaAlaLeu-----P 487
DB 1554 TGAACCTGAGTCAGACAGATCTGTCAGCCCTCTCCCTCTAAAAATTGAGGAATTAGCACT 1613
QY 487 heMetAspAspLysLeuGluAsnThrLeuLysGlyArgGlnMetLeuThrAspGluGlnL 507
DB 1614 GGCCAAAGGATCACTGAAGAATGT-CTGAAA-----CAGCATCTTTGGAAACAGAGG 1666
QY 507 ysGlyArgThrArgValLysThrIleArgAlaLeuProArgLeuPheGlyAlaProThrA 527
DB 1667 AAGGCAGAGAGCTTCTCATACCTTCTCCCAAGCCACAGATTGAACAGTCAGCTGATT 1726
QY 527 laLysAlaGluMetIleAspAlaLysValPheGlnAspIleGluGluArgProLleProP 547
DB 1727 CATCTCTAGCGGTCCTCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 1786
QY 547 roLeuPheGluProLysGly-----ArgHisThrArgLeuArg----- 560
DB 1787 CT-----GACAGTTGAGTTCTCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1837
QY 561 -----TrpThrGlyAlaAsnGluLysGluIleProGlyLeuGlySerArgP 576
DB 1838 ATGTAGCCAGGACGCTACTCATGCCAACCTCTCGTGTGTAGACCCCAAGATGGGTCCAGAT 1897
QY 576 heIleLeuProSerLeuAspProThrMetProAlaLeuAsnThrAlaPheSer----- 593
DB 1898 CAACATCAGATCAGATCAAGGTCACGTTTCACGTTCTCTCTCTCTCTCTCTCTCTCTCTCT 1957
QY 594 -----ThrGlnGlyArgAlaArgAspGluTrpAspThrMetPheLysIleProA 610
DB 1958 GAAATCTCTGAGCCCTGGAGTCTCCAGGAGACG-----AGCACCGCTATCTACTGAAACCA 2014
QY 610 snAsnTrpAsnProGlyAspGluVal 618
DB 2015 AAGATCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2040

```

RESULT 7

AK030316 2993 bp mRNA linear HTC 03-APR-2004
 LOCUS AK030316.1 GI:26326304
 DEFINITION Riken musculus 11 days pregnant adult female ovary and uterus cDNA,
 product: similar to CDNA FLJ14503 FIS. CLONE NT2RM1000252, WEAKLY
 SIMILAR TO H.SAPIENS E-MAP-115 MRNA [Homo sapiens], full insert
 sequence.

ACCESSION AK030316
 VERSION AK030316.1
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE 1
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning
 Meth. Enzymol. 303, 19-44 (1999)
 PUBMED 99279253
 REFERENCE 2
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 PUBMED 20499374
 REFERENCE 3
 Shibata, K., Itoh, M., Aizawa, K., Nagasaka, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, J., Nishi, K., Kiteunai, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

passed the following selection criteria: Hexamer frequency ORF analysis
This clone has the following problem: frame shifted.

FEATURES

source

Location/Qualifiers

```
1. 7139
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6333274"
/tissue_type="Embryo, day 9 mouse (C57BL/6 background)
otocyst8"
/lab_host="DH10B"
/note="vector: pcMV-SPORT6.1"
```

ORIGIN

Alignment Scores:

```
Pred. No.: 0.00476 Length: 7139
Score: 153.50 Matches: 131
Percent Similarity: 34.21% Conservativity: 102
Best Local Similarity: 19.24% Mismatches: 262
Query Match: 4.66% Indels: 187
DB: 3 Gaps: 28
```

US-10-736-868-2 (1-643) x BC050866 (1-7139)

```
Qy 31 AlaLeuSerSerLeuValGlySerHisGlnLysLeuProLeuAlaProSerMet 50
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1323 TCATTATCGAGAGCTGTTGGTAAAGATCACAGCCAGCGGCACCGAG----- 1376

Qy 51 GluAlaLeuGluMetGlyValGlnPheValAspAlaLeuLysGlyGlnMet 70
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1377 -----CAGTTGATGAGCACCCTTATACGAGACCACTTAATGAGGAGCGTC 1427

Qy 71 GluMetAlaLysGlyAlaPheLysThrGlnLeuGluValLeuGluLysValHisProAsp 90
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1428 -----CGCATCCAGCTGAAG-----GAC 1445

Qy 91 GlnPheAspLysTyrLysLysLeuLysValAspAspLeuAlaAlaAspAlaValMetGln 110
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1446 CACATTGATCGAACAAAGAGAGAGCGAGGAGAAAGATGAGACTGAGTATGAATACAGC 1505

Qy 111 GlnAlaGluMetAlaLysLeuGlnProLysSerGlyAsnAlaPheLysMetLeuAsn 130
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1506 GGAAGTGAGGAGAGAGAGAGAGATGACTCTCGGGAACCC---AGCTCCATTCTGAAC 1562

Qy 131 GlyAsnGlyLeuProLysSerLysLeuGluAspAlaLysLeuGluGln 150
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1563 -----CTACAGGGGAGTCAACACTGCGA-----AGGGACTTCTTGAGACTGCAG 1607

Qy 151 ArgAspMetGluAnThrAspProSerGluGlnLysAlaValMetAspLysPhe 170
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1608 -----CTGGCCAAACAGGAGCGCTCAGAGCCCTCGCGGC----- 1643

Qy 171 GlnThrGlnLysLeuProGlyLeuValAlaAsnMetIleAlaGlyLysAsnProPheLys 190
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1643 ----- 1643

Qy 191 MetProGlnGlnMetArgLysAlaGlnAlaAlaProSerSerValPheGlnAlaLeu 210
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1644 -----CAACAGCTGGAGCAGCAGCGGAGAGATGAAGAACACAGCGCGCTACTG 1697

Qy 211 AlaGlnArgAlaMetLeuGlyLysAsnAlaProValAlaGlyArgGlyGluGln 230
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1698 GCTGAGCGCCAGAAG-----CGCATCGAAGAGCAG 1727

Qy 231 ArgMetMetAsnArgVal---AspGlnArgMetGlnGlnArgGluGlnGluGlu 249
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1728 AAGAGCAAAAGCGGAGCTGGAGAGACAAAGGCGAGNAAAAGAGCTTCGGAACAG 1787

Qy 250 AspGluAspAspAspLeuGluAspGluAspValProArgArgArgSerSerAspGly 269
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1788 CAGGAGCGGGAACAGCGCGGCACTACCAAGAACAGATGCTCGGAGGAGGAGGAGG 1847
```

```
Qy 270 GluProGlnSerGluAlaGluHisGlnArgArgAspLeuAlaArgArgLeuLysSerSer 289
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1848 CGTCCGCAACATGAGCAGGAATATATAAGCGCAAAATTTGAAGAACACAGACAGAGAA 1907

Qy 290 ProArgLeuLysGluLeuLeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrGlnArg 309
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1908 AGACTCAGAGGCGCTAAAGCAAGAGCGGACTATCTGGTTTCCCTCCAGCATCAGCGG 1967

Qy 310 MetArgAspSerProLeuSerLysArgArgProLeu-----AlaMetAsn 324
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1968 CAGGAGCAGAGCGCCCTGGAGAAG---AAGCCACTGTACCATTACAGGAGGCGCATGAGT 2024

Qy 325 AspGluAspGluSerAlaPhe---ArgAlaMetGluAlaArgAlaLysLeuAspGlnLys 343
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2025 CCTAGTGAGAAAGCGCGCTGGGCCAAGAGGAGTGAAGAAACGCTCAAGACTCAACCCAGAC 2084

Qy 344 SerGlnLeuValLeuGlyLeuHisGlyPheGlyGlu-----SerAspAspAspGluAsp 361
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2085 AGTTCCACTGCCATGCT---CACAGGTTGCCAACAGGATCTCGGACCCCACTGCC 2141

Qy 362 GluGluAspGluAsnLeu----- 367
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2142 CCAAGATCAGAGTCTTTCAGCATTTAGTGGGTTTCAGCTGCAAGGACACCCCAATGCTC 2201

Qy 368 -----IleAspPro----- 370
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2202 AGACCTGTTGAGACCCCGAGATCCCGCAGCTGGTAGCTGTCTMAATCCAGGAGACCTGCC 2261

Qy 371 ---SerGluAsnSerPheArgArgAlaProLeuArgLeuSerSerGlyPheValGluLys 389
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2262 ACGCCTCCAGTCAGTACATGAGCAACCCACACAGGCGCTGCTGGTTCCAGGAG--- 2318

Qy 390 LeuLysSerAsnAspGluLeuLysSerAlaLeuAspArgLysTyrArgValAsp--- 408
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2319 -----GCTCTGAATGTGACCTCTCACCGGTCGAGATG 2351

Qy 409 -----AspValGluLysTyrLeuAlaProLysProMetGluPheAsnPro 423
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2352 CCAGCCAGACACTCGGATCCCACTCAGAAAACCTCTCTCCAGGAGAAATGAGAAG 2411

Qy 424 LysProGlnProGlyTyrPhe-----AlaProArgLysLysProThrArg 438
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2412 TTTGACAGAAAGCTCTTGGTTTACGACAGGAAGAGACATTCCACCAAGGTGCTCAAGA 2471

Qy 439 ProArgLysMetLeuProLeuLeu----- 446
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2472 ACAACTTCTATATCCCGCAGCACTAGCCAGAAAGAAATTCCTCGCAATGGCACTGCTCTG 2531

Qy 447 -----IleGlySerAspPro-----LysValGlnGluLysLysArgArg---His 460
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2532 GGCCCGAGACTTGGATCTGAGCCCATCAGAGCAAGCAACCTGATCTGCGCAGGACAGAG 2591

Qy 461 ProSerThrGluThrLysLysLysLysGluSerArgValLeuThrAsnLeuLysAsnAsn 480
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2592 CCAGTCTGAGAGTTCCTGCGACGCAAGCAGTGGCAGTTCCTCCAGCTCCAGCACT 2651

Qy 481 ProSerLeuAlaAlaLeuPheMetAspLysLeuGluAsnThrLeuLysGlyArgGln 500
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2652 CCCAGC-----TCCAGCCCGCAGCTCCCAAGAGGCTCTCAA 2687

Qy 501 MetLeuThrAspGluGlnLysGlyArgThrArgValLysThrIleArgAlaLeuProArg 520
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2688 -----CCTGGCTCCCAAGCAGGATCTAGTACGGGTCCAGAGTGGGGCAACAGTAAG 2741

Qy 521 LeuPheGlyAlaProThrAlaLysAlaGlu-----MetIleAspAlaLysValPheGln 538
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2742 TCCGAAGGATCACCCGTGCTCCCGCATGACCTTCCAAAGGTGAACCCAGAAAGATCCAGA 2801

Qy 539 Aspile-----GluGluArgProLysPhePheGluProLysGlyArgHis 556
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 2802 GACATCACGCGCCCGCTCGGCCAGCTGATCTGCGGCATTCGCAAAAGAAATACGAGAA 2861
```

Qy 557 ThrArgLeuArgTrpThrGlyAlaAsnGluLysGluLeuProGlyLeuGlySerArg--- 575
 Db 2862 CTCGCAATTGAAGAAACAAACCGCCCTTGAAGAAAGTGAAGTACTTCTCTCCAGC 2921
 Qy 575 ----- 575
 Db 2922 GAGGAGTCGAGAGCAGTGGAGGAAGAGGAGAGTGGAGAGTGGAGACACATGACGGG 2981

Qy 576 -----PheIleLeuProSerLeuAspProThrMetProAlaLeuAsnThrAla 591
 Db 2982 ACGGTGGCTGTCAAGTACATCCAGACTAATACCCAGGAGTCCAGGAGCAATGAG 3041
 Qy 592 PheSerThrGlnGlyArgAlaArgPheGluTrpAspThrMetPheLysIleProAsnAsn 611
 Db 3042 CA-GTACACATGGGATGGTCGGGACACATGGCTGGAAACTTGGCATCGGACACCTT 3100

Qy 612 TTP 612
 Db 3101 TGG 3103

RESULT 9
 BC050868 7139 bp mRNA linear HTC 25-MAR-2004
 LOCUS IMAGE:6334359), containing frame-shift errors.
 DEFINITION Mus musculus TRAF2 and NCK interacting kinase, mRNA (CDNA clone

ACCESSION BC050868
 VERSION BC050868.1 GI:30047833
 KEYWORDS HTC.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 7139)
 AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Datchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Uedin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullen, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Pahey, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

TITLE
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
 PUBMED 12477932
 REFERENCE 2 (bases 1 to 7139)
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (09-APR-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 COMMENT Contact: MGC help desk
 Email: cgapsb@mail.nih.gov
 Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman, Ph.D.
 cDNA Library Preparation: ResGen, Invitrogen Corp
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)
 DNA Sequencing by: Genome Sequence Centre,
 BC Cancer Agency, Vancouver, BC, Canada
 info@bcgsc.bc.ca
 Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield,

Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao, Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin, Teika Olson, Diana Palmquist, Anca Petrescu, Anna Liisa Prabh, Parvaneh Saedi, JR Santos, Angeliq Schnerch, Ursula Skalska, Duane Smalls, Jeff Stott, Miranda Tsai, George Yang, Jacqueline Schein, Asim Siddiqui, Rob Holt, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAP Plate: 109 Row: d Column: 16
 This clone has the following problem: frame shifted.

FEATURES
 Location/Qualifiers
 1..7139
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="IMAGE:6334359"
 /tissue_type="Embryo, day 9 mouse (C57BL/6 background)"
 otocysts"
 /clone_lib="NIH_MGC_130"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6.1"

ORIGIN
 Alignment Scores:
 Pred. No.: 0.00476 Length: 7139
 Score: 153.50 Matches: 131
 Percent Similarity: 34.21% Conservativeness: 102
 Best Local Similarity: 19.24% Mismatches: 262
 Query Match: 4.66% Indels: 187
 DB: 3 Gaps: 28

US-10-736-868-2 (1-643) x BC050868 (1-7139)
 Qy 31 AlaLeuLeuSerSerLeuValGlySerHisGlnLysLeuProLeuAlaProSerMet 50
 Db 1323 TCATTATCGAGAGCTGCTTGGTAAAGAAATCACAGCCAGCGCCAGCAGGAG 1376
 Qy 51 GluAlaLeuGluLeuMetGlyValGlnPheValAspAlaLeuLysGlyGlnMet 70
 Db 1377 -----CAGTTGATGAAGAGCACCCTTATACAGAGCAACCTAATGAGGAGCGTC 1427
 Qy 71 GluMetAlaLysGlyAlaPheLysThrGlnLeuGluValLeuLysValHisProAsp 90
 Db 1428 -----CGCATCCAGCTGAAG-----GAC 1445
 Qy 91 GlnPheAspLysTyrLysLysLeuValAspAspLeuAlaAlaAspAlaValMetGln 110
 Db 1446 CACATTGATCGAACAAGAGAGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1505
 Qy 111 GlnAlaGluMetAlaLysLeuGlnProLysSerGlyAsnAlaPheIleAspMetLeuAsn 130
 Db 1506 GGAAGTGAAGGAAG 1562
 Qy 131 GlyAsnGlyIleProLysGlySerSerIleArgGlyLeuGluAspAlaIleArgThrGln 150
 Db 1563 -----CTACACAGGGAGTCAACTGCGA-----AGGAGCTTCTGAGAGTGCAG 1607
 Qy 151 ArgAspMetGluAsnThrAspProSerGluGlnIleAlaLysAlaValMetAspLysPhe 170
 Db 1608 -----CTGGCCACAAGAGAGCGCTCAGAGCGCCCTCGCGCGC----- 1643
 Qy 171 GlnThrGlnIleLeuProGlyLeuValAlaAsnMetIleAlaGlyLysAsnProPheLys 190
 Db 1643 ----- 1643
 Qy 191 MetProGlnGlnMetArgLysAlaGlnAlaAlaProSerSerValPheGlnGlnAlaLeu 210
 Db 1644 -----CAACAGCTGGAGCAGCAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1697
 Qy 211 AlaGlnArgAlaMetLeuGlyLysAsnAlaProValAlaGlyArgGlyGluGluGln 230
 |||:|||||

Db 1698 GCTGAGCGCCAGAG-----CGCATCGAAGACGAG 1727
 Qy 231 ArgMetMetMetAsnArgVal---AspGlnArgMetGlnArgGluLeuGlnGlu 249
 Db 1728 AAGGAGCAAAAGCGCGAGCTGCTGAGGAGCAACAAGGCGAGAAAGAGCTTCGGAACAG 1787
 Qy 250 AspGluAspAspAspAspLeuGluAspGluAspValProArgArgArgSerSerAspGly 269
 Db 1788 CAGGAGCGGGAACAGCCCGCAGCTACGAAAGACAGATGCGTGGGAGGAGGAGGAGG 1847
 Qy 270 GluProGlnSerGluAlaGluHisGlnArgArgAspLeuAlaArgArgLeuLysSerSer 289
 Db 1848 CGTCCCGAACATGAGCAGGATATAGCGCAACAAATTGGNAGNACAGACAGACAGAA 1907
 Qy 290 ProArgLeuLysGluLeuLeuGlnAsnAlaGluValGlnSerLeuLeuSerTyrGlnArg 309
 Db 1908 AGACTGCAGAGGCGAGCTAAAGCAAGAGCGGAGCTATCTGTTCCCTCCAGCATCAGCGG 1967
 Qy 310 MetArgAspSerProLeuSerLysArgArgProLeu-----AlaMetAsn 324
 Db 1968 CAGGAGCAGAGCGCCCTGGGAAG---AAGCCACTGTACCATTAAGAGGAGGCGATGAGT 2024
 Qy 325 AspGluAspGluSerAlaPhe---ArgAlaMetGluAlaArgAlaLysLeuAspGlnLys 343
 Db 2025 CCTAGTCAGAGCGGCTGGCCCAAGGAGGAGTAGAAGACGCTCAAGACTCAACCGACAG 2084
 Qy 344 SerGlnLeuValLeuGlyLeuHisGlyPheGlyGlu-----SerAspAspAspGluAsp 361
 Db 2085 AGTTCACGTCGCATGCT---CACAAAGTTGCCAAAGGATCTCGGACCCCAACCTGCC 2141
 Qy 362 GluGluAspGluAsnLeu-----367
 Db 2142 CCAAGATCAGAGTCTTTCAGCATTTAGTGGGTTTCAGCTGCAAGGACACCCCAAGTCTC 2201
 Qy 368 -----IleAspPro-----370
 Db 2202 AGACTGTGACCCCGAGATCCCGCAGCTGGTAGCTGCTCAATCCAGGAGACCTGCGCTG 2261
 Qy 371 ---SerGluAsnSerPheArgAlaProLeuArgLeuSerSerGlyPheValGluLys 389
 Db 2262 ACGGCTCCAGTCAGTACATGAGCAACCCCAAGGCGCTGCTGGGTTCCAGGAG---2318
 Qy 390 LeuLysSerAsnAspGluLeuLysSerAlaLeuAspArgIleLysTyrArgValAsp---408
 Db 2319 -----GCTCTGAATGTGACCTCTCACCGGTCGAGATG 2351
 Qy 409 -----AspValGluLysTyrLeuAlaProLysProMetGluPheAsnPro 423
 Db 2352 CCAGCCAGAACTCGGATCCACCTCAGAAACCTCTCTCCCCACAGAAATTGAGAAG 2411
 Qy 424 LysProGlnProGlyTyrPhe-----AlaProArgLysIleProThrArg 438
 Db 2412 TTTGACAGAGCTCTTGTTACGACAGGAAGACATTCACCAAAAGTGCCTCAAGA 2471
 Qy 439 ProArgLysMetLeuProLeuLeu-----446
 Db 2472 ACAACTTCTATCCCGCAGCATCCGACAGAAAGAAATTCCTCCGCAATGGCAGTGTCTG 2531
 Qy 447 -----IleGlySerAspPro-----LysValGlnGluGluIleArgArg---His 460
 Db 2532 GGCCCCAGACTTGGATCTCAGCCCATCAGACAGCAACCTGTATCTGCGCAGGACAGAG 2591
 Qy 461 ProSerThrGluTyrLysIleAlaLysGluSerArgValLeuThrAsnLeuLysAsnAsn 480
 Db 2592 CCAGTCTCGAGAGTTCCTCGCAGCGGACAGCAGTGCAGTTCCTCCAGCTCCAGCACT 2651
 Qy 481 ProSerLeuAlaLeuPheMetAspLysLysLeuGluAsnThrLeuLysGlyArgGln 500
 Db 2652 CCCAGC-----TCCAGCCCGCAGCTCCCAAGGAGGCTCTCAA 2687
 Qy 501 MetLeuThrAspGluGlnLysGlyArgThrArgValLysThrIleArgAlaLeuProArg 520
 Db 2688 -----CCTGGCTCCCAAGCAGGATCTAGTGAGCGGTCAGAGTGGCGGCAACAGTAAG 2741

Qy 521 LeuPheGlyAlaProThrAlaLysAlaGlu-----MetIleAspAlaLysValPheGln 538
 Db 2742 TCCGAAGGATCACCCCTGTCTCCCCATGAGCTTCCAAGGTGAACACGAGAAGATCCAGA 2801
 Qy 539 AspIle-----GluGluArgProIleProProLeuPhePheGluProLysGlyArgHis 556
 Db 2802 GACATCACAGGCCAGTGGCCAGCTGATCTGACGGCATTAGCCAAAGATTTACGAGAA 2861
 Qy 557 ThrArgLeuArgPrpThrGlyAlaAsnGluLysGluIleProGlyLeuGlySerArg---575
 Db 2862 CTCGCGATTGAAGAAACAAACCCGCCCTGGAAGAAAGTGACTGATTACTTCTCTCCAGC 2921
 Qy 575 -----575
 Db 2922 GAGGAGTCCGAGAGCAGTGGAGGAAGAGGAGATGGAGAGAGTGAGACACATGACGGG 2981
 Qy 576 -----PheIleLeuProSerLeuAspProThrMetProAlaLeuAsnThrAla 591
 Db 2982 ACGTGGCTGTAGTCACATACCCAGCTAATACCCACCGGAGCTCCAGGGAACATGAG 3041
 Qy 592 PheSerThrGlnGlyArgAlaArgAspGluTyrAspThrMetPheLysIleProAsnAsn 611
 Db 3042 CA-GTACAACATGGGATGTCGCGACACATGGGCTGGAAACTTCGCATCGCGACACCTT 3100
 Qy 612 Ttp 612
 Db 3101 TGG 3103
 RESULT 10
 AY421013
 LOCUS
 DEFINITION Mus musculus ACINUS gene, VIRTUAL TRANSCRIPT, partial sequence,
 genomic survey sequence.
 ACCESSION AY421013.1 GI:39776970
 VERSION
 KEYWORDS GSS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 REFERENCE 1 (bases 1 to 3918)
 AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
 Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
 Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
 Adams,M.D. and Cargill,M.
 TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
 gene trios
 JOURNAL Science 302 (5652), 1960-1963 (2003)
 PUBMED 14671302
 REFERENCE 2 (bases 1 to 3918)
 AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
 Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
 Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
 Adams,M.D. and Cargill,M.
 TITLE Direct Submission
 JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
 COMMENT This sequence as made by sequencing genomic exons and ordering them
 based on alignment.
 FEATURES
 source 1..3918
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 <1..>3918
 /genes="ACINUS"
 /locus_tag="HCMT415"
 gene
 ORIGIN
 Alignment Scores: 0.00247 Length: 3918
 Pred. No.: 152.50 Matches: 132
 Score:

Percent Similarity: 36.02% Conservativity: 96
 Best Local Similarity: 20.85% Mismatches: 230
 Query Match: 4.63% Indels: 176
 DB: 9 Gaps: 24

US-10-736-868-2 (1-643) x AY421013 (1-3918)

QY 23 ProSerThrSerAspAlaProGlyAlaLeuLeuSer---SerLeuValGlyLysSerHis 41
 DB 157 CCGAGTACCTCAGCAAGATGCGGAGCTGGAGGAGTGACTCTGGACGGGAAGCCTCTT 216
 QY 42 GlnLysLeuProLeuAlaProSerMetGluAlaLeuGluLeuMetGlyValGlnPheVal 61
 DB 217 CAGGCTCTCGCGGTGACCGACCTGAAGCCGCACTGGAGCAGCGAGGC----- 264
 QY 62 AspAlaLeuLeuLysLysGlyGln-----MetGluMetAlaLysGly----- 75
 DB 265 -----CTAGCCAGAGCGGCGAGAGAGTGCCTGGTCAAGCGCTCAAGGGGCTCTA 318
 QY 76 -----AlaPheLysThrGlnLeuGlu 82
 DB 319 ATGCTAGAAAATTACAAAAACACTCAACACCCCATGCTGCATTCCAGCCAAATTCACAG 378
 QY 83 ValLeuGluLysValHisProAspGlnPheAspLysTyLysLysLeuLysValAspAsp 102
 DB 379 ATCGGGGAGGAGATGAGCCAGAACAGCTTCATAAAGCAGTACCTGGAAAGACGACGAG 438
 QY 103 LeuAlaAlaAspAlaValMetGlnGlnAla---GluMetAlaLysLeuGlnProLysSer 121
 DB 439 CTGCTGAGCGCAGCGCTGGACGGGAGAGCTCGGAGGCTGGAGGCTGGAGAGACCTCA 498
 QY 122 GlyAsnAlaPheIleAspMetLeuAsnGlyAsnGlyIle-----Pro 135
 DB 499 GCTGAGTCGGAGGACGAGATGACCCATCTCGAGGGAGTGCTTCCCTGCTGCTCCTGAT 558
 QY 136 IleGlySerSerIleArgGlyLeuGluAspAlaIleArgThrGlnArgAspMetGluAsn 155
 DB 559 TTTCAGACGACGCTGAACAGACGACGAGCTGGAGCTCAGCACACATTCACCCCGAAGAGC 618
 QY 156 ThrAspProSerGluGlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeu 175
 DB 619 TCCTCTTTTCTGAAGAA----- 636
 QY 176 ProGlyLeuValAlaAsnMetIleAlaGlyLysAsnProPheLysMetProGlnGlnMet 195
 DB 637 -----AAGGAGAAATCTGATGATGAGAACCG----- 663
 QY 196 ArgLysAlaGlnAlaAlaProSerSerValPheGlnGlnAlaLeuAlaGlnAlaMet 215
 DB 664 AGGAAGGGGAAAGACGATCATCTCAGAGTT---AGACAGGCAAAATCTAAACTCCCTGAG 720
 QY 216 LeuGlyLysAsnAlaProValAlaGlyArgGlyGluGlnArgMetMetMetAsn 235
 DB 721 TACAGTCAGCTGAGGAGGAGAGATCAAGAGACACCATCTAGAAACCTCAGGGTC 780
 QY 236 ArgValAspGlnArgMet-----GlnGlnArgGluLeuGlnGluAspGluAsp 252
 DB 781 AGAGCAGATCGAAATTTAAATAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGAT 840
 QY 253 AspAspAspLeuGluAspValProArgArgSerArgSerAspGlyGluProGln 272
 DB 841 GATGATGATGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 867
 QY 273 SerGluAlaGluHisGlnArgArgAspLeuAlaArgArgLeuLysSerSerProArgLeu 292
 DB 868 GATAGGCACAGAAATCTAGA----- 888
 QY 293 LysGluLeuLeuGlnAsnAlaGluValGlnSerLeuLeuSerTyGlnArgMetArgAsp 312
 DB 889 -----GAGGCTGAGGCACCAACCTGAGCAGCTTTGAGGATGAGAGAGGG 933
 QY 313 SerProLeuSerLysArgArgProLeuAlaMetAsnAspGluAspGluSerAlaPheArg 332
 DB 933 ----- 933

DB 934 GAAGAGAGGACCCAGAGCAAAACCCAGAAAGGTGGTGGATGAGAAACCCCTAAACATCAGA 993
 QY 333 AlaMetGluAlaArgAlaLysLeuAsp---GlnLysSerGlnLeuValLeuGlyLeuHisG1 352
 DB 994 TCCAGGNA---AAGGTGAGTTAGAGAAAGAGGAGGCGAGTTACTA----- 1036
 QY 352 yPheGlyGluSerAspAspGluAspGluGluAspGluAsnLeuLeuLeuAspProSerG1 372
 DB 1037 -----GATCCCAAGAAAGACCTAGAAAGTCAATCGCCACAGACAG 1077
 QY 372 uAsnSerPheArgArgAlaProLeuArgLeuSerSerGlyPheValGluLysLeuLysSe 392
 DB 1078 CAACAAGAAAGAGACACAAATTTGTTCTCTCC-----CTCAAGAA 1119
 QY 392 xAsnAspGluLeuLysSerAlaLeuAspArgIleLysTyArgValAspAspValGluLyl 412
 DB 1120 -GAAATGAGTAAAGTCTTCA-----CAAGTTTAGAGGAAATAATCCCA 1163
 QY 412 sTyLeuAlaProLysProMetGluPheAsnProLysProGlnProGlyTyPheAlaPr 432
 DB 1164 GTCTCCCTCCCACTCGCTTCCTGAAGACCTAGAGAGGCGCTGTGTGTGTGCAGCC 1223
 QY 432 oArgLysIleProThrArgProArgLysMetLeuProLeuLeuLeu-----GlySerAs 450
 DB 1224 AGAAGATAGTCACTGAGTGAAGAGAAACACCCCACTTACTTACCAAGSAGCGTCAATC 1283
 QY 450 pPro-----LysValGlnGluGluIleArg----- 458
 DB 1284 TCCACCACTCATATACACTCCAGGAGAGATGGAGCCAGTGAAGCCCGCAGACCCCCC 1343
 QY 459 -----ArgHisProSerThrG1 464
 DB 1344 TGCTCTCATTGCTGCT 1403
 QY 464 uTrpLysIleAlaLysGluSerArgValLeuThrAsnLeuLysAsnAsnProSerLeuAl 484
 DB 1404 TCCTCACCTCGCAGCTGGTGAAGAGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1452
 QY 484 aAlaLeuPheMetAspAspLysLeuGluAsnThrLeuLysGlyArgGlnMetLeuThrAs 504
 DB 1453 -----ACAGACACCAAGCAGAGTCT----- 1473
 QY 504 pGluGlnLysGlyArgThrArgValLysThrIleArgAlaLeuProArgLeuPheGlyAl 524
 DB 1474 ---CCAGCAGGAGAGTGTCT 1529
 QY 524 aProThrAlaLysAlaGluMetIleAspAlaLysValPheGlnAspIleGluGluArgPr 544
 DB 1530 GCCTGAATGCTCCGCCAG-----GAGGTGTGTTGAAAGTGAAGCGTGAAGAGTC 1577
 QY 544 o-IleProProLeu-----PhePheGluProLysGlyArgHisThrArgL 559
 DB 1578 TGCTCCCT 1637
 QY 559 eu-----ArgTrpThrGlyAlaAsnGluLysGluIleProGlyLeuGlySerA 575
 DB 1638 TATGAAGAAACAGCTCTTTGGAACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1697
 QY 575 rgPheIleLeuProSerLeuAspProThrMetPro 586
 DB 1698 AGAGTACAGTGGGAGCAGTCACTCATCTCT 1732

RESULT 11

AY414528
 LOCUS 2934 bp DNA linear GSS 17-DEC-2003
 DEFINITION Homo sapiens RNF20 gene, VIRTUAL TRANSCRIPT, partial sequence,
 genomic survey sequence.
 ACCESSION AY414528
 VERSION AY414528.1 GI:39770490
 KEYWORDS GSS.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
              1. (bases 1 to 2934)
              Clark, A.G., Gnanowksi, S., Nielson, R., Thomas, P., Kejarawal, A.,
              Todd, M.A., Tanenbaum, D.M., Civeello, D.R., Lu, F., Murphy, B.,
              Ferrera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
              Adams, M.D. and Cargill, M.
              Inferring nonneutral evolution from human-chimp-mouse orthologous
              gene trios
              Science 302 (5652), 1960-1963 (2003)
              14671302
REFERENCE
AUTHORS      2. (bases 1 to 2934)
              Clark, A.G., Gnanowksi, S., Nielson, R., Thomas, P., Kejarawal, A.,
              Todd, M.A., Tanenbaum, D.M., Civeello, D.R., Lu, F., Murphy, B.,
              Ferrera, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,
              Adams, M.D. and Cargill, M.
              Direct Submission
              Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
              Rockville, MD 20850, USA
              This sequence was made by sequencing genomic exons and ordering
              them based on alignment.
FEATURES
source       Location/Qualifiers
              1. .2934
              /organism="Homo sapiens"
              /mol_type="genomic DNA"
              /db_xref="taxon:9606"
              <1..>2934
              /gene="RNF20"
              /locus_tag="HCM5254"
gene
Alignment Scores:
Pred. No.:      0.00181      Length:      2934
Score:          152.00      Matches:    108
Percent Similarity: 39.19%      Conservative: 104
Best Local Similarity: 19.96%      Mismatches: 197
Query Match:    4.62%      Indels:    133
DB:             9          Gaps:       22
US-10-736-868-2 (1-643) x AY14528 (1-2934)
QY 58 ValGlnPheValAspAlaLeuLysGlyGlnMetGluMetAla----- 73
DB 1174 GTCTGTATATGAGACCTACAGTTGAAGACACACTGGAGGTCGGACCTGCTT 1233
QY 74 -----LysGlyAlaPheLysThrGlnLeuGluValLeuGluLysValHisProAsp 90
DB 1234 CATGGCACGAGAGACCCACCACGACGAGTTGAGCTTATGAGCGA-----GAT 1284
QY 91 GlnPheAspLysTyrlsLysLeuLysValAspLeuAlaAspAlaValMetGln 110
DB 1285 GAGGTTAGTCTTCATAGAGCTGAGGACTGAAGTAATTCAGCTAGAGATACATTGGCC 1344
QY 111 GlnAla-----GluMetAlaLysLeuGlnProLysSerGlyAsnAlaPheIle 126
DB 1345 CAGTCCGCAAGGAGTATGAATGCTGAGGATAGATTGAGCAGACCTTGCT----- 1398
QY 127 AspMetLeuAsnGlyAsnGlyIleProIleGlySerSerIleArgGlyLeuGluAspAla 146
DB 1399 -----GCCAATGAACAAGCAGGCGCTATATAACAGGAGGAGATGCGCCACCTCATCAGTAGC 1452
QY 147 IleArgThr-----GlnArgAspMetGluAsnThrAspProSerGluGlnIleAla 163
DB 1453 CTCAGAAATCAATACCACTGAAAGGGGAGGTCCTGAGATATAGCGGAATTTGAGA 1512
QY 164 LysAlaValMetAspLysPheGlnThrGlnIleLeuProGly-----LeuValAlaAsn 181
DB 1513 GAAGCCAGCTGACCTGAACAAGACACGCTGCTGAGTGGTAGTGCCTCTCGCAGTCC 1572
QY 182 MetIleAlaGlyLysAsnProPheLysMetProGlnGlnMetArgLysAlaGlnAlaAla 201
DB 1573 CAGTCTAGTACTAGGACCCGAGGATGAGCTGCGGAGCTTAAACCCAGCATTTCTGAGGAC 1632
QY 202 ProSerValPheGlnAlaLeuAlaGlnArgAlaMetGluLysAsnAlaPro 221

```

```

DB 1633 TTATCTCTCC-----CAGTCCTCAGCTTCAAAGGCATCTCAGAGGATGCCAATGAA 1683
QY 222 ValAlaGlyGlyArgGlyGluGlnArgMetMetMetAsnArgValAspGlnArgMet 241
DB 1684 ATCAAGTCTAAACCGGATGAAGAAGAACGAGAACGAGAAAGGAGGAGAGAGAGGAA 1743
QY 242 GlnGlnArgGlu-----LeuGlnGlu 248
DB 1744 CGAAGAAGAGAACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1803
QY 249 GluAspGluAspAspAspAspLeuGluAspValProArgArgArgSerSerAsp 268
DB 1804 TCAGAAAAGAGAGAGATTTCTGTAAGGATTAAGAG-----AAAGCCAAACATGATGAT 1857
QY 269 GlyGluProGlnSerGluAlaGlu-----HisGlnArgArgAspLeuAlaArgArg 285
DB 1858 GGACGG-----AAAAGGAAGCAGAAATTATCAACAATTTGAAGATTGAACCTCAAGAAGGCA 1914
QY 286 LeuLysSerSerProArgLeuLysGluLeuLeu----- 296
DB 1915 CAGGAGAGCCAAAGAGGAGATGAACACTATTGCTGGATATGTACCGTTCTGCCCCCAAGGAA 1974
QY 297 -----GlnAsnAlaGluValGlnSer 303
DB 1975 CAGAGAGACAAAGTTTCAGCTGATGGCAGCTGAGAGAAGTCTTAAGCGAGTTGGAAGAT 2034
QY 304 LeuLeuSerTyrlsGlnArgMetArgAsp-----SerProLeuSerLysArgArgProLeuAla 322
DB 2035 CTA-----AGGCAAGACTCAAGGATCTGGAAGATAAAGAGAGAAAGAGAGAAAGAA 2088
QY 323 MetAsnAspGluAspGlu-----SerAlaPheArgAlaMetGlu----- 335
DB 2089 ATGGCTGATGAGGATCCCTTGAGGAAGATCCCGGCGAGTGGAGGAGGAGATAGATACCTA 2148
QY 336 -----AlaArgAlaLysLeuAspGlnLysSerGlnLysValLeuGlyLeuHis 351
DB 2149 CAGAAGAAGCTAGCCATGCGCAAGCAGGAGAGAGAGAGAGCA---CTCCTCTCTGAATGGAT 2205
QY 352 GlyPheGlyLysSerAspAspGluAspGluGluAsp-----GluAsnLeu 367
DB 2206 GTCACAGGCGCAGCCCTTTGAAGACATGCGAGGACAAATATATCCGTTTGTGTCGACCAATTG 2265
QY 368 IleAspProSerGluAsnSerPheArgAlaProLeuArgLeuSerSerGlyPheVal 387
DB 2266 CCGGAGAGAGATGATCAAAATTTCAAGCTCATGTCAGAGCGTATCAAGTCCAATCAGATC 2325
QY 388 GluLysLeu-----LysSerAsnAspGluLeuLysSerAlaLeuAspArgIleLysTyr 405
DB 2326 CATAGTTGCTTAAAGAGAGAGAGAGAGCTGGCAGACCAGGTGTTGACTCTGAGACT 2385
QY 406 ArgValAsp----- 408
DB 2386 CAGGTTGATGCCCGCAGCTACAGGTAGTAAGGAAACTGGAAGAGAGAGAGCATCTGTTACAG 2445
QY 409 -----AspValGluLysTyrlsLeuAlaProLysProMetGluPheAsn 422
DB 2446 AGCAACATTGGCAGAGGAGAGAGCTGGGTCTTAGGACCCAGCCTTAGAGATGAAT 2505
QY 423 ProLysProGlnProGlyTyrlsPheAlaProArgLysIleProThrArgProArgLysMet 442
DB 2506 AAACGC----- 2520
QY 443 LeuProLeuLeuGlySerAspProLysValGlnGluGluLeuArg-ArgHisPro-- 461
DB 2521 GAGGCGAGCCAGCTTGCAGATGACCTCAAGACCAACTGGAGTTGGCTGAGAGAGAGCTA 2580
QY 462 -----SerThrGluTrpLysIleAlaLysGluSerArgValLeuThrAsnLe 477
DB 2581 CATGATTTTCAGGATGAGATCGTGGAGAACAGCTGTTACCAAGAGAGAGCATGTTCAAT 2640
QY 477 uLysAsnAsnProSerLeuAlaLeuPheMetAspLysLeuGluAsnThrLeuLys 497

```


Db	2641	TTCAACGAGCCCGAGGAGCATCTCTAGA-CTTCGCGAGGAGCTGAGACCAAAAGAA	2699
Qy	497	s 497	
Db	2700	A 2700	
RESULT 12			
CL960352			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
ORIGIN			
Alignment Scores:			
Pred. No.:			
Score:			
Percent Similarity:			
Best Local Similarity:			
Query Match:			
DB:			
US-10-736-868-2 (1-643) x CL960352 (1-3012)			
Qy	24	SerThrSerAspAlaProGlyAlaLeu-----LeuSerSerLeu	36
Db	589	TCGACGAATGCTGCTATCGGATGCTTCTAGAGGAGAGCAGTCTTGATTTGTCCAAGTTG	648
Qy	37	ValGlyLysSerHisGlnLysLeuProLeuAlaProSerMetGluAlaLeuGlu-----	54
Db	649	GTGGATGAGATATATGATAAAGCTCAAGGAATGAAGCGGTGTCGGCGACTTCGGTCCGG	708
Qy	55	-----LeuMetClyValGlnPheValAspAlaLeu	64
Db	709	AGTTCTGTGCTCATTTGGCGCAGAGATGATGTATGGGCAATCTAGTCCGGATGCTGAT	768
Qy	65	IleLysLysGlyGlnMetGluMetAla-----	73
Db	769	GTGTTGGAAGATGAATCAGAGACGGCTCTTTTGGTGGGAGGTGAGAGATTGGAAGTG	828
Qy	74	-----LysGlyAlaPheLysThrGlnLeuGluValLeuGluLysValHis	88
Db	829	ATTCCCTCTGAAGATGCGGTGGACCTCTCAGTATCGGCTAGAACTGCTGAGAAAGAAATCCAC	888

Qy	405	TyrArgValAspAspVal	-----	410
		:		
Db	1903	-----GTTGATATGCTACCAGCGGTTGAGCTACAGTTTCATGGCCACCAATCATGCTAAAT	1956	
Qy	411	-----GluIysTyrLeuAlaProIlysProMetGluPheAsnProIlysProGlnPro	427	
		: : : : : :		
Db	1957	CCTCTGCCAATAGATCAATAAGCGGAAGCTTTTGCAATTGTGCAAAAGCAACAGACCT	2016	
Qy	428	GlyTyrPheAlaPro--ArgLysIle-----ProThrArgProArgLys	441	
		: : : : :		
Db	2017	GCATATTATGGAGCTGTGGAGGAAGAAAGTCCTGCTGTGTCGCCCAAGATGCCACTTAAG	2076	
Qy	442	MetLeuProLeuLeuIleGlySerAspProIlysValGlnGluGluIleArgArgHisPro	461	
		: : : : :		
Db	2077	ATG-----GATCCTGATCTTGACTATGAGGTT-----GATAGT	2109	
Qy	462	SerThrGluThrIysIleAlaLysGluSerArgValLeuThrAsnLeu--LysAsnAsn	480	
		: : : : :		
Db	2110	GATGATGAATGGGAAGAGGAGATCTGTGTGAGAGCCTTTCCGACTCGGAGAAGGACAAT	2169	
Qy	481	ProSerLeuAlaAlaLeuPheMetAspAspIlysLeuGluAsnThrLeuLysGlyArgGln	500	
		: : : : :		
Db	2170	-----GATCAAGTTATGGAGGAGGATTCAG-----	2196	
Qy	501	MetLeuThrAspGluGln	506	
		: : : : : :		
Db	2197	---ATCACAGATGAAGAG	2211	

RESULT 13					
CR679256					
LOCUS	1128 bp	mrna	linear	HTC 19-AUG-2004	
DEFINITION	Tetraodon nigroviridis full-length cDNA.				
ACCESSION	CR679256				
VERSION	CR679256.1	GI:51177163			
KEYWORDS	HTC; cDNA; full-length; Tetraodon nigroviridis.				
SOURCE	Tetraodon nigroviridis				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.				
REFERENCE	1 (bases 1 to 1128)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (10-AUG-2004) Genoscope - Centre National de Sequencage -				
	: 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE				
	(E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)				
COMMENT	The sequences are based on single pass reads.				
	More information available at				
	http://www.genoscope.cns.fr/tetraodon .				

ORIGIN

Alignment Scores:	
Pred. No.:	0.00623
Score:	150.50
Percent Similarity:	36.6%
Best Local Similarity:	22.7%
Query Match:	4.57%
DB:	3
Length:	1128
Matches:	90
Conservative:	55
Mismatches:	156
Indels:	95
Gaps:	17

US-10-736-868-2 (1-643) x CR679256 (1-1128)

Qy	104	AlaAlaAspAlaValMetGlnGlnAlaGluMetAlaLysLeuGlnProLysSerGlyAsn	123
		: : : : : : : : : :	
Db	92	GCCATGATGCCATCAAGAAGAGATGCAGATGCTCAAGCTCGACAGGAACACTGCC---	148
Qy	124	AlaPheIleAspMetLeuAsnGlyAsnGlyIleProIleGlySerSerIleArgGlyLeu	143

Db 1061 CCGCGC-----CGGCACACACAGCAGCACTCAGCAG 1093

RESULT 14
AK030449
LOCUS
DEFINITION
AK030449 3431 bp mRNA linear HTC 03-APR-2004
Mus musculus adult male pituitary gland cDNA, RIKEN full-length
enriched library, clone:5330413K22 product:unknown EST, full insert
sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
AK030449.1 GI:26081823
HTC; CAP trapper.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636

2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159

3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861

4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 3431)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, P., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,

COMMENT
CDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/.

FEATURES
source
1..3431
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM DB:5330413K22"
/db_xref="taxon:10090"
/clone="5330413K22"
/sex="male"
/tissue_type="pituitary gland"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"

misc_feature
1..3431
/note="unknown EST (GB|AI593306, evidence: BLASTN, 98%,
match=441)"

ORIGIN
Alignment Scores:
Pred. No.: 0.0032 Length: 3431
Score: 150.50 Matches: 134
Percent Similarity: 35.92% Conservative: 93
Best Local Similarity: 21.20% Mismatches: 230
Query Match: 4.57% Indels: 176
DB: 3 Gaps: 25
US-10-736-868-2 (1-643) x AK030449 (1-3431)
QY 24 SerThrSerAspAlaProGlyAlaLeuLeuSer---SerLeuValGlySerHisGln 42
Db 4 AGTACCTCAGCAAGATGGGGGAGCTGGAGGAGGTGACTCTGGACGGGAAGCCTCTTCAG 63
QY 43 LysLeuProLeuAlaProSerMetGluAlaLeuGluLeuMetGlyValGlnPheValAsp 62
Db 64 GCTCTGGGGGAGCCGACCTGGAAGCCGCACTGGAGCAGCGAGGC----- 108
QY 63 AlaLeuLeuLysLysGlyGln-----MetGluMetAlaLysGly----- 75
Db 109 ---CTAGCCAGAGCGGGCAGAGAGTGCCCTGTCAAGCGCTCAAGAGGGCTCTAATG 165
QY 76 -----AlaPheLysThrGlnLeuGluVal 83
Db 166 CTAGAAAATTTACAAAACACTCAACCCCATCTGTCATTCCAGCCAAATTCACAGATC 225
QY 84 LeuGluLysValHisProAspGlnPheAspLysTyrLysLysLeuLysValAspAspLeu 103
Db 226 GGGGAGGAAATGAGCCACAGACAGCTTCTATAAGCAGATGATCTGGAAGAGCAGCAGGAGCTA 285
QY 104 AlaAlaAspAlaValMetGlnGlnAla---GluMetAlaLysLeuGlnProLysSerGly 122
Db 286 CTTAGGCAGCCCTGGAGACGGGAAGCTCGGAGAGCTCGGGAACCTTGAAGAAGCCCTCAGCT 345
QY 123 AsnAlaPheLeAspMetLeuAsnGlyAsnGlyTle-----Profile 136
Db 346 GAGTCGGAGGACGAGATGACCCATCTCTGAGGAGTGGCTTCCTGCTGCTCTCTGATTTT 405
QY 137 GlySerSerLeuArgGlyLeuGluAspAlaIleArgThrGlnArgAspMetGluAsnThr 156
Db 406 CAGAGCAGCTTGAACACAGCAGCAGCTGGAGCTCAGCACATTCACCCCGAAGAGCTCC 465
QY 157 AspProSerGlnGlnIleAlaLysAlaValMetAspLysPheGlnThrGlnIleLeuPro 176
Db 466 TCTTTTCTGAAGAA----- 480
QY 177 GlyLeuValAlaAsnMetIleAlaGlyLysAsnProPheLysMetProGlnGlnMetArg 196

Search completed: July 18, 2005, 16:09:04
Job time : 5650 secs

THIS PAGE BLANK (USPTO)